



MONTCAS, PHASE 2
Criterion-Referenced Test
Alternate Assessment
(CRT-Alternate)

2005
TECHNICAL MANUAL

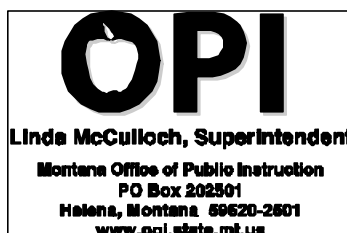


TABLE OF CONTENTS

SECTION I: ASSESSMENT DEVELOPMENT	5
CHAPTER 1—BACKGROUND AND OVERVIEW	5
PURPOSE OF THIS MANUAL	5
PURPOSE OF THE CRT-ALTERNATE	5
PARTICIPATION GUIDELINES	7
ORGANIZATION OF THIS MANUAL	8
CHAPTER 2—OVERVIEW OF TEST DESIGN	9
CRT-ALTERNATE	9
ASSESSMENT TYPE	9
CRT-ALTERNATE ITEMS	9
CHAPTER 3—TEST DEVELOPMENT PROCESS	11
ITEM AND ACTIVITY DEVELOPMENT	11
DEVELOPMENT OF THE READING AND MATHEMATICS EXPANDED BENCHMARKS	11
CRT-ALTERNATE ITEM DEVELOPMENT PROCESS OVERVIEW	13
REVISIONS MADE TO THE SPRING 2005 ASSESSMENTS	14
ITEM/ACTIVITY EDITING	15
CHAPTER 4—DESIGN OF THE READING ASSESSMENT	17
READING BLUEPRINT	17
CHAPTER 5—DESIGN OF THE MATHEMATICS ASSESSMENT	18
MATHEMATICS BLUEPRINT	18
CHAPTER 6—TEST FORMAT	19
TEST DESIGN	19
TEST ACTIVITY MATERIALS	21
SECTION II: TEST ADMINISTRATION	22
CHAPTER 7—TEST ADMINISTRATION	22
RESPONSIBILITY FOR ADMINISTRATION	22
PROCEDURES	22
ADMINISTRATOR TRAINING	24
PARTICIPATION REQUIREMENTS	24
TEST SCHEDULING	25
SECTION III: DEVELOPMENT AND REPORTING OF SCORES	27
CHAPTER 8—SCORING	27
SCORING THE ASSESSMENT	27
USING SCAFFOLDING TO GATHER STUDENT PERFORMANCE INFORMATION	27
SCORING RUBRIC	29
SCORING RULES	31
MACHINE-SCORED ITEMS	31
SCANNING QUALITY CONTROL	32
ELECTRONIC DATA FILES	33
CHAPTER 9—ITEM ANALYSES	34
DIFFICULTY INDICES	34
ITEM-TEST CORRELATIONS (ITEM DISCRIMINATION)	35
SUMMARY OF ITEM ANALYSIS RESULTS	35
DIFFERENTIAL ITEM FUNCTIONING	37
CHAPTER 10—RELIABILITY	38
RELIABILITY	39
CHAPTER 11—SCALING	41
TRANSLATING RAW SCORES TO SCALED SCORES AND PERFORMANCE LEVELS	41
CHAPTER 12—REPORTING	44
CHAPTER 13—VALIDITY SUMMARY	45

SECTION IV: REFERENCES	48
SECTION V: APPENDICES	49
APPENDIX A: ADVISORY COMMITTEE MEMBERS, TEST DEVELOPEMNT AND REVISION WORKSHOP PARTICIPANTS, AND BETA TEST PARTICIPANTS	50
APPENDIX B: TECHNICAL ADVISORY COMMITTEE	52
APPENDIX C: STANDARD SETTING REPORT AND EVALUATION SUMMARIES	53
APPENDIX D: CRT-ALTERNATE PERFORMANCE LEVEL DESCRIPTORS, SCALED SCORES AND RAW SCORES	82
APPENDIX E: CRT-ALTERNATE RELEASED PERFORMANCE INDICATORS	91
APPENDIX F: REPORT SHELLS	100
APPENDIX G: REPORTING DECISION RULES	108

SECTION I: ASSESSMENT DEVELOPMENT

CHAPTER 1—BACKGROUND AND OVERVIEW

PURPOSE OF THIS MANUAL

The purpose of this manual is to document the technical aspect of the 2005 MontCAS, Phase 2 Criterion-Referenced Test Alternate Assessment (CRT-Alternate). In the spring of 2005, students in grades 4, 8, and 10 participated in the administration of the CRT-Alternate; during this administration reading and mathematics were assessed. This represents the second year of the CRT-Alternate program which will expand during the next two years to include additional grades (3, 5, 6 and 7) and science grades 4, 8 and 10. This report provides information about the technical quality of those assessments, including a description of the processes used to develop, administer, and score the tests and to analyze the test results.

Historically, while some parts of a technical report may have been used by educated laypersons, the intended audience was experts in psychometrics and educational research. This edition of the CRT-Alternate technical report is a first attempt to make the information contained herein more accessible to educated lay people by providing richer descriptions of general categories of information. In making some of the information more accessible we have purposefully preserved the depth of technical information that has historically been provided in our technical manuals. The reader will find that some of the discussion and tables continue to require a working knowledge of measurement concepts such as “reliability” and “validity,” and statistical concepts such as “correlation” and “central tendency.” To fully understand some data, the reader will also have to possess basic familiarity with advanced topics in measurement and statistics.

PURPOSE OF THE CRT-ALTERNATE

The Individuals with Disabilities Education Act (IDEA) requires that students with disabilities be included in each state’s system of accountability and that students with disabilities have

access to the general curriculum. The No Child Left Behind Act (NCLB) also speaks to the inclusion of all children in a state's accountability system by requiring states to report student achievement for all students as well as for groups of students on a disaggregated basis. These federal laws reflect an ongoing concern about equity: All students should be academically challenged and taught to high standards. It is also necessary that all students be involved in the educational accountability system.

To ensure the participation of all students in the state's accountability system, Montana has developed the Criterion-Referenced Test Alternate Assessment (CRT-Alternate). The CRT-Alternate is an evidence-based test that is aligned with Montana's content standards through expanded benchmarks and measures student performance based on alternate achievement standards. It is expected that only those IDEA-eligible students with the most significant cognitive disabilities will participate in the CRT-Alternate.

On April 5, 2002, the Montana Office of Public Instruction (OPI) entered into a compliance agreement with the U.S. Department of Education that required Montana to implement a number of actions by April 5, 2005, to bring the state into compliance with the provisions of the following federal laws: Title 1 of the Elementary and Secondary Education Act (ESEA) of 1994, P.L. 103-382 and the No Child Left Behind Act (NCLB) of 2001. Montana received federal appropriations to develop an appropriate assessment. The criterion-referenced test Alternate Assessment (CRT-Alternate) was developed in accordance with the compliance agreement and federal laws.

The CRT-Alternate assessments are based on, and aligned to, Montana's Content Standards and Expanded Benchmarks in Reading and Mathematics. Montana educators worked with OPI and its contractor, Measured Progress, in the development and review (content and bias) of these tests to assess how well students have learned the Montana Content Standards and Expanded Benchmarks for their grade. The United States Department of Education (USDOE) approved the CRT-Alternate assessments in reading and mathematics for grades 3–8 and 10 by school year 2005-2006 and in science at one grade in each of three grade spans (e.g., 4, 8 and 10) by school year 2007–2008.

The CRT-Alternate is a new assessment. This was the second year of implementation. After the first year, extensive revisions were made based on feedback from teachers who administered the assessment. Alternate assessments have only been in place since 2000. The field is still in the learning stages as to appropriate ways to address reliability and validity for alternate assessments.

PARTICIPATION GUIDELINES

The decision as to how a student with disabilities will participate in the state's accountability system is made by the student's Individualized Education Program (IEP) team. When considering whether students with disabilities should participate in the CRT-Alternate, the IEP team should address each of the questions in the chart that follows:

For each of the statements below, answer YES or NO		
Does the student have an active IEP and receive services under the Individuals with Disabilities Education Act (IDEA)?	YES	NO
Do the student's demonstrated cognitive abilities and adaptive behavior require substantial adjustments to the general curriculum?	YES	NO
Do the student's learning objectives and expected outcomes focus on functional application of skills, as illustrated in the student's IEP's annual goals and short-term objectives?	YES	NO
Does the student require direct and extensive instruction to acquire, maintain, generalize and transfer new skills?	YES	NO

- If you answer "NO" to any of the above questions, the student must participate in the regular CRT.
- If all answers are "YES," the student is eligible to take the alternate and is considered to be a student with a significant cognitive disability.

The decision to determine a student's eligibility to participate in the CRT-Alternate may not be based on

- excessive or extended absence;
- disability category;
- social, cultural or economic difference;
- the amount of time receiving special education services; or
- academic achievement significantly lower than his or her same age peers.

Because the regular CRT provides full access to the vast majority of students, it is expected that only approximately 100 students per grade will participate in the CRT-Alternate.

In accordance with 34 CFR 200.13 Adequate Yearly Progress in General, there is a 1% cap applied to the number of proficient and advanced scores based on the alternate assessment that may be included in AYP calculations at both the state and district levels.

ORGANIZATION OF THIS MANUAL

The organization of this manual is based on the conceptual flow of an assessment's life span: it begins with the initial test specification and addresses all the intermediate steps that lead to final score reporting. Section I covers the development of the CRT-Alternate tests. It consists of six chapters covering general design, the test development process, the specific designs of the reading and mathematics assessments, and the test format. Section II consists of a single chapter describing the administration of the tests. Section III contains six chapters covering scoring, item analysis, reliability, scaling, reporting, and validity. Also included are two additional sections: Section IV contains references and Section V contains the appendices.

CHAPTER 2—OVERVIEW OF TEST DESIGN

CRT-ALTERNATE

CRT-Alternate test items are directly linked to **Montana's Content Standards and Expanded Benchmarks** (see page 11 for more information about the Expanded Benchmarks). The content standards are the basis for the reporting categories developed for each subject area and are used to help guide the development of test items. No other content or process is subject to statewide assessment. An item may address part, all, or several of the benchmarks within a standard or standards.

ASSESSMENT TYPE

The CRT-Alternate assessment is a point-in-time test that looks at how students perform in relation to performance indicators which have been expanded from the Montana reading and mathematics standards and benchmarks. Each content area consists of one age-appropriate activity that consists of twenty to thirty items in which the teachers are given a script, written directions and scaffolding levels. Teachers have the option of using the suggested activity or creating their own activity. If the teacher decides to create their own activity, an outline of the activity used must be submitted. Students are encouraged to engage in the activity and show performance on the items through appropriate prompting by the teacher who administers the test activity. The teacher who administers the test activity scores the student on each item through observation using a five-point scoring rubric.

The test activity requires evidence to be collected based on the products that were created during the course of the assessment. Templates were provided for all evidence that was required.

CRT-ALTERNATE ITEMS

Each item of the CRT-Alternate consists of the following:

- Materials needed to administer the item,
- Set-up instructions and script for the teacher to follow if using the suggested test activity,
- Scaffolding script for the suggested test activity,

- The correct student response,
- The performance indicator, (The performance indicator is what the question is measuring. The performance indicator comes from the Montana Standards and Expanded Benchmarks.)
- Activity steps to follow for teachers creating their own activity.

See Chapter 6 for the test format.

CHAPTER 3—TEST DEVELOPMENT PROCESS

ITEM AND ACTIVITY DEVELOPMENT

The CRT-Alternate was developed as a collaborative project between Measured Progress and the Montana Office of Public Instruction divisions of Assessment, Special Education, and Educational Opportunity and Equity.

An advisory committee, representing perspectives of parents, teachers, administrators, and faculty in higher education, provided input during the development of this assessment. In addition, teacher work groups were formed at several points in the development and revision process. Math and reading item development work groups, composed of general and special education teachers, were formed. These teachers developed test activities that are the basis of the performance tasks for this assessment. A third group of special education teachers and administrators participated in the beta testing of this assessment, providing valuable feedback about the test design.

DEVELOPMENT OF THE READING AND MATHEMATICS EXPANDED BENCHMARKS

The expanded benchmarks were developed for students with significant cognitive disabilities not working at the same level as their age level counterparts. The expanded benchmarks were developed using Montana's standards and benchmarks for reading and mathematics. Measured Progress curriculum and special education specialists developed a draft of the expanded benchmarks. The OPI, beta test teachers, the Advisory Committee, and the development and revision workshop participants all provided input and recommendations for changes to the original draft. Using these recommendations Measured Progress revised the expanded benchmarks. This document was further revised to include grade span expectations per new federal legislation. It is expanded from end of grade 4, end of grade 8 and end of grade 12—upon graduation to foundational skills. These are not grade level specific, due to the wide diversity of students in this population. This document was used to develop the assessment performance indicators. The chart on the next page shows how the document is organized and gives an example for each content area. The Montana Content Standards and Expanded Benchmarks are not included in this manual because of the length

of each document. They are located on the OPI Web site at www.opi.state.mt.us and the Measured Progress Web site at www.measuredprogress.org.

Montana CRT – Alternate Standards and Expanded Benchmarks

Terminology		
Term/Description	Example	
Content Area	Mathematics	Reading
Standard Learning outcome expected for all students throughout all grades	Standard 2: Students demonstrate understanding of and ability to use Numbers and Operations.	Standard 2: Students apply a range of skills and strategies to read.
Essence of the Standard A statement of the standard separating the essential components	Number concepts, concepts of operations, computing and estimating	Interpret print and nonprint information
Benchmark Grade Level Expectation (GLE) Expectation for typical students described for each grade level	2.2, Grade 4: Students will use the number system by counting, grouping, and applying place value concepts.	2.6, Grade 8: Students will develop vocabulary through the use of context clues, analysis of word parts, auditory clues, and references sources (e.g., dictionary, thesaurus, and glossary).
Expanded Benchmark Benchmark skill or concept expanded from the typical GLE to a basic level	2.2.1: The student will demonstrate an understanding of whole numbers.	2.6.2: The student will use word/pictures/symbols/objects to communicate.
Performance Indicator Expanded benchmark expressed in a measurable and observable statement of a specific performance	2.2.1.2: The student will demonstrate the concept of one (e.g., "Hit the switch one time"; "Give me one").	2.6.2.1: The student will identify a word/picture/symbol/object used to name a familiar place.
Prompt The script for the directions the test administrator will deliver to the student, calling for the specific behavior	Item 4: "These are counters. We are going to use these in our activity. Show me one counter."	Item 4: "Show me the word/picture/symbol/object that means 'library.'"

TOTAL NUMBERS OF ITEMS DEVELOPED BY GRADE AND CONTENT

GRADE	READING	MATH
4	22	28
8	24	32
10	27	31

CRT-ALTERNATE ITEM DEVELOPMENT PROCESS OVERVIEW

An overview of the test development process for the CRT-Alternate program, including conducting the beta test (November 2003), follows.

DEVELOPMENT PROCESS OVERVIEW

DEVELOPMENT STEP	PROCEDURE OF THE STEP
Development and Revision of Expanded Benchmarks for reading and mathematics (Aug. 2003-Oct. 2004)	<ul style="list-style-type: none"> Measured Progress curriculum and special education specialists developed a draft of the expanded benchmarks. The OPI reviewed it. Beta test teachers provided input. The Advisory Committee and revision and development workshop participants provided recommendations. The expanded benchmarks were revised to include grade span expectations per new federal legislation.
Development workshops and assessment development (Aug. 2003)	<p>Measured Progress curriculum and special education specialists and the OPI</p> <ul style="list-style-type: none"> provided item development training to Montana participants; facilitated the development of the item ideas by the participants; used the items and activities that were developed at the workshops to finish developing the assessments.
Editorial review of items (Oct. 2003)	<p>All items were reviewed by members of the Measured Progress Publications staff to ensure -</p> <ul style="list-style-type: none"> clarity and unambiguousness of items; correct grammar, punctuation, usage, and spelling; technical quality with respect to stems, options, and scoring guides; compliance with OPI sensitivity standards and style guidelines.
Beta test (Nov. 2003)	<ul style="list-style-type: none"> Approximately 20 students participated in the beta test for grades 4, 8 and 10. Beta test teachers tested a student on 1 content area, and sent feedback to Measured Progress on the assessment items and activity. Beta test participants gave additional feedback in a

	<p>conference call.</p> <ul style="list-style-type: none"> • The Advisory Committee reviewed all grades and contents and provided feedback via a form and conference call.
Revisions after beta test (Nov. 2003 – Jan. 2004)	<ul style="list-style-type: none"> • Using the feedback from the beta test teachers and the Advisory Committee, the OPI and Measured Progress revised the assessment. • Level 1 scaffolding script was added to every item on the test that is scored using all five levels of the rubric.
First administration of the assessment (Feb. – Mar. 2004)	<ul style="list-style-type: none"> • Approximately 100 students per grade were assessed. • Teachers sent feedback on every item using a template provided with the assessment. • The OPI developed an online survey for teachers to provide additional feedback on the assessment.
Revision workshop (Oct. 2004)	<ul style="list-style-type: none"> • Measured Progress curriculum and special education specialists and the OPI facilitated the workshop. • Approximately 12 teachers participated. • Teachers revised the assessment items and activities. • Measured Progress and the OPI continued to revise the assessment after the workshop and sent it to Publications for another editorial review.
Second test administration (Feb. 2005 – Mar. 2005)	<ul style="list-style-type: none"> • Approximately 120 students were tested per grade.

REVISIONS MADE TO THE SPRING 2005 ASSESSMENTS

Using feedback from teachers who administered the CRT-Alternate in the spring of 2004, Montana special education and general education teachers, the OPI, and Measured Progress revised the following in the assessments:

- Level 1 scaffolding language was added to the *Suggested Activity, Teacher will:* column.

This was added to give teachers a clearer direction on how to at scaffold this level.

- The *Materials for the Suggested Activity* column was added. This column lists the materials needed for each item, as well as communication support strategies. This column was added to prepare teachers on what materials are needed to administer each

item and for students to respond to each item. It also gives teachers ideas for student communication supports.

- Ancillary materials and training CDs were developed and sent to teachers administering the assessment.
- Optional breaks were added to giving teachers a clearer idea of when to give the student a break in the test activity.
- Item language was revised for clarity and consistency.
- Items were added and deleted to help cover all standards evenly across all grades (3-8 and 10).
- The scoring rule for halting the assessment was changed from *“Score every item until the student scores in level 1 or 0 for five consecutive items. Halt the administration if the student scores in level 1 or 0 for five consecutive items. Leave the remaining items blank.”* to *“Score every item until the student scores at level 0 for three consecutive items. Stop the administration of the assessment at this point. On the following assessment session, re-administer the final three items on which the student scored a 0. If the student receives a level 0 on three consecutive items again, halt the administration of the assessment and leave the remaining items blank.”* Three examples were given for this new rule. This was based on in depth discussion with the TAC and recommendations that they made.

ITEM/ACTIVITY EDITING

Editors reviewed and edited the items and test activities to ensure uniform style (based on *The Chicago Manual of Style*) and adherence to sound testing principles. These principles included the stipulation that items and the test activities:

- were correct with regard to grammar, punctuation, usage, and spelling;

- were written in a clear, concise style;
- were measuring the performance indicator;
- had materials that were appropriate;
- contained unambiguous explanations for teachers as to what was required of the student;
- were written at a reading level that would allow the student to demonstrate his or her knowledge of the tested subject matter regardless of reading ability;
- exhibited high technical quality regarding psychometric characteristics;
- had appropriate scaffolding script for teachers; and
- were free of potentially insensitive content.

CHAPTER 4—DESIGN OF THE READING ASSESSMENT

READING BLUEPRINT

As indicated earlier, the framework for reading was based on Montana's Standards and Expanded Benchmarks, which identifies five **content standards** that apply specifically to reading and reading comprehension. Those content standards are:

- **Reading Standard 1:** Students construct meaning as they comprehend, interpret, and respond to what they read.
- **Reading Standard 2:** Students apply a range of skills and strategies to read.
- **Reading Standard 3:** Students set goals, monitor, and evaluate their reading progress. (This standard is not measurable in a statewide assessment.)
- **Reading Standard 4:** Students select, read, and respond to print and nonprint material for a variety of purposes.
- **Reading Standard 5:** Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences.

The chart below shows the standards measured at each grade level.

	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5
Grade 4	9	9	*	3	1
Grade 8	10	10	*	2	2
Grade 10	13	7	*	3	4

*Standard 3 is not measurable in a statewide assessment.

Note — Decisions for the test blueprints were made by looking at concepts across all grades and where the focus/introduction of the other concepts occurs in general education so that students who are included in general education classrooms and activities are also working on similar expanded skills. Standards 1 and 2 for both math and reading are measured at every grade level and the other standards are measured evenly across grade spans (elementary 3-5, middle 6-8 and high school 10).

CHAPTER 5—DESIGN OF THE MATHEMATICS ASSESSMENT

MATHEMATICS BLUEPRINT

The mathematics framework was based on Montana’s Mathematics Content Standards and Expanded Benchmarks, which identifies seven **content standards**, as shown below:

- **Mathematics Standard 1:** Problem Solving
- **Mathematics Standard 2:** Numbers and operations
- **Mathematics Standard 3:** Algebra
- **Mathematics Standard 4:** Geometry
- **Mathematics Standard 5:** Measurement
- **Mathematics Standard 6:** Data Analysis, Statistics, and Probability
- **Mathematics Standard 7:** Patterns, Relations, and Functions

The chart below shows the standards measured at each grade level.

	STANDARD 1	STANDARD 2	STANDARD 3	STANDARD 4	STANDARD 5	STANDARD 6	STANDARD 7
Grade 4	9	8	0	0	0	13	4
Grade 8	7	8	4	0	5	11	0
Grade 10	5	13	7	4	0	0	3

Note — Decisions for the test blueprints were made by looking at concepts across all grades and where the focus/introduction of the other concepts occurs in general education so that students who are included in general education classrooms and activities are also working on similar expanded skills. Standards 1 and 2 for both math and reading are measured at every grade level and the other standards are measured evenly across grade spans (elementary 3-5, middle 6-8 and high school 10).

CHAPTER 6—TEST FORMAT

TEST DESIGN

The CRT-Alternate is composed of two test activities: reading and mathematics. The teacher has the option of administering the suggested activity or a similar test activity created by the teacher. If a teacher designed activity was given, an outline of the activity was to be submitted with the test booklet and answer sheet. Each test activity consists of 20 to 30 items and at least one piece of student evidence (work). Since only one test was developed, every student takes the same form of the test. The test stays the same each year, with the exception of the second year. Revisions to the test were made using teacher feedback and a revision workshop. Only the Performance Indicators, which come from the Montana Reading and Mathematics Standards and Expanded Benchmarks, are released every year on the OPI and Measured Progress Web sites. The 2005 released Performance Indicators are located in Appendix E.


The first page of the math and reading sections of the test booklet lists the following:

- content standards
- brief explanation of the suggested test activity
- parameters of the task
- materials provided and other materials that are needed

The pages that follow in the math and reading sections of the test booklet consist of the following five columns for each item:

Materials for the Suggested Activity	Suggested Activity Teacher will:	Student Work Student will:	Performance Indicators Use Scoring Guide TRANSFER SCORES TO STUDENT RESPONSE BOOKLET	Activity Steps ONLY NEEDED IF NOT USING SUGGESTED ACTIVITY Teacher will:
The materials that are needed for each item and suggested student communication supports and strategies that may be helpful for some students are described in this column. Most materials can be found in the Material Kit, but some materials teachers need to supply.	<p>This column contains information about how to display task materials and prepare the student for the question. A script for the teacher appears in bold and italicized print, suggests language that can be used to present the item. The script is intended as a guide only, and should be adapted by the teacher as needed.</p> <p>Information on how to scaffold levels 3, 2, and 1 of the rubric for items that are scored at levels 4 through 0 is also provided in this column.</p>	The correct student response and/or an explanation of how the student should be responding is provided in this column.	The performance indicator that is assessed by each item is identified in this column. The performance indicators come from the Montana Standards and Expanded Benchmarks.	Gives the teacher instructions on what to ask the student if the teacher is not administering the suggested activity. If the teacher is administering the suggested activity, this step does not need to be used.

Evidence and Evidence Template(s)

Each of the test activities requires that evidence be collected based on the products that are created during the course of the assessment. A magnifying glass  in the “Student Work, Student will:” column of the test booklet indicates when evidence must be collected. Templates are provided in the CRT-Alternate Test Booklet for all evidence that is required. Teachers have the option of selecting the presentation that best matches the student’s abilities and skills:

- Written work by the student (e.g., the student collects data and fills out a bar chart with a marker)
- Pictures of student output (e.g., the student arranges objects to form an answer to a question about the sequence of events in a story and a picture captures the arrangement)
- Picture symbols pasted on the template or a scanned/photocopied image of the template that the student arranges and that he/she wants to keep
- Computer printout of student's keyed responses
- Teacher-recorded responses (e.g., the teacher fills out a T-table based on the yes/no answers from a student using a BIGmack switch or eye gaze)
- Anecdotal record describing student's actions supplied by the observer (e.g., the observer notes that the student smiled when shown a picture of his/her favorite character in a story)

The evidence templates are used to record student responses to an item when asked. Adapted evidence templates are provided in the materials kit and on the materials CD. The template may need further modifications based on the student's needs.

Last Page of the test booklet

The last page of the test booklet contains a list questions for the teacher to answer after the administration of the reading and mathematics test activities.

TEST ACTIVITY MATERIALS

Ancillary materials were developed for the math and reading test activities by Gail McGregor from the Rural Institute for Disabilities, University of Montana at Missoula. These materials include picture response choices for questions that are structured as multiple choice items and picture symbols that can be used to support communication for students who require augmentative communication supports. Material kits contained laminated, color copies of each material needed for the activities. Also included in the material kit was a CD that contained all the electronic files in case a teacher needed to change materials to meet the needs of their student, and a teacher training CD.

SECTION II: TEST ADMINISTRATION

CHAPTER 7—TEST ADMINISTRATION

RESPONSIBILITY FOR ADMINISTRATION

The special education teacher or someone who is certified and has worked extensively with the student and is trained in the assessment procedures administers the assessment. The test administrator may find it helpful to ask another person in the school to assist with the administration.

These additional persons may include but are not limited to the following:

- parent
- general education teacher
- paraprofessional
- special service provider (speech/language therapist, psychologist, occupational or physical therapist, etc.)
- school counselor
- principal
- other educational professional

PROCEDURES

Teachers administering the CRT-Alternate were sent a training CD with an audio PowerPoint to train them on implementing the test. The following are the procedures teachers were given to prepare to administer the assessment:

- View training CD and participate in question/answer sessions.
- Receive the secure *CRT-Alternate Test Booklet* from the test coordinator.
- Receive hard copy of the test activity materials, CD with test activity materials, and training CD from Gail McGregor at the Rural Institute of Disabilities, University of

Montana at Missoula. Teachers may have needed to further adapt materials to meet the need of the students taking assessment. Guidelines and examples for adapting materials were given in the “Materials” section of the test booklet and in the *CRT-Alternate Administration Manual*.

- Download the *CRT-Alternate Administration Manual* and Scoring Rubric from the OPI or the Measured Progress Web site.
- Read the *CRT-Alternate Administration Manual* to become familiar with the administration and scoring directions.
- Read the *CRT-Alternate Test Booklet* to become familiar with the test activity steps and performance indicators.
- Determine the test activity that will be used—either the suggested test activity or a similar test activity that the teacher creates.
- Consider how the student will access and respond to the test activity. Determine the adaptations and supports that the student will need.
- Check to ensure that all of the materials and resources needed to complete the test activity are available. For example: The grade 8 reading activity asks the student to locate the library and to identify the librarian. The reference or book area in the classroom may be substituted for the library, and someone who helps students pick a book (i.e., teacher) may be substituted for the librarian.
- Provide the assistive technologies that the student needs to access the materials and respond to the test activities.
- Schedule the assessment administration session for a time and place that are optimal for student effort and focus.

School Test Coordinators were instructed to read the *Test Coordinator’s Manual* prior to testing, and to be familiar with the instructions given in the *Test Administrator’s Manual* and the *CRT-Alternate Administration Manual*. The *Test Coordinator’s Manual* and the *CRT-Alternate Administration Manual* provided each school with checklists to help prepare for testing. The checklists outlined tasks to be performed before, during, and after test administration. Along with providing these checklists, the *Test Coordinator’s Manual* and the *CRT-Alternate Administration Manual* outlined the nature of the testing material being sent to each school, how to inventory the material, how to track it during administration, and how to return the material once testing was complete. It also contained information about including

or excluding students. The *CRT-Alternate Administration Manual* included a checklist for the test administrators to prepare themselves, their classrooms, and their students for the administration of the test and how to return the assessment.

ADMINISTRATOR TRAINING

In addition to distributing the *Test Coordinator's Manual* and *CRT-Alternate Administration Manual*, OPI and Measured Progress conducted preadministration workshops on February 8, 2005 (one MetNet and one videostream) to train and inform school personnel about the CRT and CRT-Alternate. Training materials and the PowerPoint presentation were posted on the OPI's Web site. In addition, teacher training CDs were sent to every teacher administering the CRT-Alternate.

PARTICIPATION REQUIREMENTS

All students were expected to participate; however, scores of students in the following categories were excluded from the calculation of averages:

- Foreign exchange students
- Students not enrolled in an accredited Montana school (for example: home schooled student)
- Students enrolled in a private accredited school
- Students enrolled in a private nonaccredited school
- Students enrolled in a private nonaccredited Title 1 school
- Students enrolled part-time (less than 180 hours) taking a mathematics or reading course
- First year in US LEP students **were required** to participate in the math assessment only; they were excluded from the reading assessment.

SUMMARY OF ELIGIBILITY FOR EXCLUSION FROM THE CRT AND CRT-ALTERNATE

EXCLUDED FROM AVERAGES	MUST PARTICIPATE	MAY PARTICIPATE
Foreign exchange students	YES	
Students not enrolled in an accredited Montana school		YES
Students enrolled in a private accredited school	YES	
Students enrolled in a private nonaccredited school		YES
Students enrolled in a private nonaccredited Title I school		YES
Students enrolled part-time (less than 180 hrs.) taking a mathematics or reading course		YES
Reading: First year in US LEP students		YES
Mathematics: First year in US LEP students	YES	

Information about the exclusion was coded in by staff after testing was completed. The *Test Coordinator's Manual* and *Test Administrator's Manual* provided directions on coding. Please refer to Appendix G: Reporting Decision Rules regarding reporting exclusions.

TEST SCHEDULING

The CRT-Alternates were given during the spring: **reading** and **mathematics** were administered to grades 4, 8 and 10 during a six-week window (February 14 – March 29, 2005). Schools were able to schedule testing sessions at any time during this period. This window, longer than that for the CRT, allows teachers administering the CRT-Alternate extra time to prepare and adapt test activity materials needed for testing.

The CRT-Alternate is an untimed assessment. Teachers administering the assessment were instructed to watch students for indications that a break may be needed. Breaks were inserted in the test booklet. Teachers could choose to stop at the breaks inserted or at other points in the assessment.

SECTION III:

DEVELOPMENT AND REPORTING OF SCORES

CHAPTER 8—SCORING

SCORING THE ASSESSMENT

Teachers administered the assessment to a student one-on-one or with the help of another administrator. The teacher scores every item as it is administered using the rubric and a process called scaffolding.

USING SCAFFOLDING TO GATHER STUDENT PERFORMANCE INFORMATION

Scaffolding is a process of providing the student the support needed to respond to the questions in the test activity. During daily instruction, many strategies are used frequently to ensure that students experience success. For example, if a student is unable to make a correct choice from a display of four pictures, the teacher reduces the complexity of the test activity by removing one of the choices. Scaffolding serves this same function and is provided so that students will experience success in completing the test activities. An important result of scaffolding is that it helps students demonstrate their knowledge and skills. These skills can be described and measured, resulting in an accurate picture of what students can do.

The scoring system in the CRT-Alternate is built on increasing amounts of scaffolding, provided only when the student does not respond or responds incorrectly. This approach is sometimes described as a “least to most” prompt hierarchy.

Each test activity begins with items that introduce the subject and materials that will be used in the test activity. These items are scored as either a 4 (student responds accurately and with no assistance) or a 0 (student does not respond or actively resists). Items that are scored at a level 4 or 0 may also be found further into the activity when new materials are being introduced.

After these items are scored, each subsequent item within the test activity will be scored on a five-point scale 4–0, with 4 representing a correct, independent response and 1 representing a correct response that has been completely guided by the teacher. A score of 0 will be used when the student does not respond or actively resists participation in the test activity. See the scoring rubric on page 30.

A script is provided for scaffolding for each of the suggested test activities. It describes the prompts that can be used to scaffold the student to a level 3, level 2 and level 1. It may be used verbatim or modified by the teacher to meet the needs of the student. For each test item, level 1 prompting is full support from the teacher to guide the student to the correct response. Depending on the student and the test item, this may involve physically guiding the student to the correct response, or some other form of support that ensures that the student responds correctly. If the teacher decides to create a test activity rather than use the suggested one, the scaffolding script may be used as a guide.

It is critical that the test administrator deliver each item in a way that allows the student the opportunity to score at level 4. That is, assume that the student can respond independently to each item, even if that is not the usual instructional practice. Follow the guidelines to observe the student demonstrating the performance required and allow adequate wait time for the student to process the information and respond without assistance. Do not repeat the questions multiple times. Then, if the student does not respond or responds incorrectly, scaffold the student to level 3—“student responds accurately when teacher clarifies, highlights important information, or reduces the range of options to three.” Again, give the student adequate wait time. If the student does not respond or responds incorrectly, scaffold to level 2—“student responds accurately when teacher provides basic yes/no questions or forced choices between two options.” If the student still does not respond with the desired behavior, scaffold to level 1—“student is guided to correct response by teacher (e.g., modeling the correct response or providing full physical assistance).” If the student resists participating for an item, the test administrator will indicate a 0—“student does not respond or actively resists.”

Scaffolding is based on the amount of information the student needs to reach the correct response. If the student can respond independently (4), no further information is needed by

the student. If the student does not respond accurately or independently, more information is given about the item and the choices are reduced (3) [see script in the CRT-Alternate Test Booklet]. This funneling toward the correct response continues as the student needs more assistance—by providing specific information about the item and a forced choice between two options (2) [see script in the CRT-Alternate Test Booklet], and finally, to guiding the student to the correct response (1) [see script in the CRT-Alternate Test Booklet]. In this way, the student is not expected to “get it” or “not get it,” as in most on-demand assessments. The CRT-Alternate considers the level of assistance that students need to demonstrate their knowledge and skills and thus provides more precise information about student performance and achievement. This system is sensitive to small increments of change in student performance, an important consideration in describing the learning outcomes of students with severe disabilities.

This process must be used systematically with each item identified for scoring within the test activity. The intent is to give the student every opportunity to perform independently on each item.

Scaffolding examples are given in the *CRT-Alternate Administration Manual*.

SCORING RUBRIC

Each test activity begins with introductory items. Only rubric levels of 4 and 0 will be used to score these introductory items. Items that are scored at a level 4 and 0 may also be found further into the assessment when new materials are being introduced. All five levels of the rubric are used to score remaining items. Teachers administering the assessment are encouraged to have the rubric available as a reference when giving the test. The five levels of the rubric are on the following page.

Montana Alternate Assessment Scoring Guide

Performance (independence and accuracy)

Used to score every item during the structured observation test activity.

4	3	2	1	0
Student responds accurately and with no assistance.	Student responds accurately when teacher clarifies, highlights important information or reduces the range of options to three.	Student responds accurately when teacher provides basic yes/no questions or forced choices between two options.	Student is guided to correct response by teacher (e.g., modeling the correct response or providing full physical assistance).	Student does not respond or actively resists.

SCORING RULES

The instructions and examples illustrate the following rules for scoring:

- Begin with the introductory items and score 4 or 0.
- Use the full scale of 4, 3, 2, 1, and 0 to score the test activity items. Start with level 4 and work systematically through the scaffolding system for every performance indicator, as necessary based on the student's response.
- Allow for appropriate wait time as you scaffold through each level of the scoring rubric.
- Do not repeat questions or directions numerous times.
- Visual, verbal, gestural, and physical cues are allowed in each level except 4.
- Record only one score for each item.
- Score 0 only if the student does not respond or actively resists.
- Halt the administration if the student is showing a pattern of resisting, is becoming fatigued or is not participating in any way, and resume testing at another time.
- Score every item until the student scores at level 0 for three consecutive items. Stop the administration of the assessment at this point. At the following assessment session, readminister the final three items on which the student scored a 0. If the student receives a level 0 on three consecutive items again, halt the administration of the assessment and leave the remaining items blank.

MACHINE-SCORED ITEMS

Once the 2005 test booklets had been logged in, identified with appropriate scannable, preprinted school information sheets, examined for extraneous materials, and batched, they were moved into the scanning area. For all student response booklets (and other forms that required imaging/scanning) this was the last step in the processing loop in which the documents themselves were handled.

At that point, 100 percent of the student response documents and other scannable information necessary to produce the required reports had been captured and converted into an electronic format, including all student identification and demographics, and digital image clips of short-answer and constructed-response student responses. The digital image clip information allowed Measured Progress to replicate student responses on the readers'

monitors just as they had appeared on the originals. From that point on, the entire process—data processing, data analysis, and reporting—was accomplished without further reference to the originals.

The first step in that conversion was the removal of the booklet bindings so that the individual pages could pass through the scanners one at a time. Once cut, the sheets were put back in their proper boxes and placed in storage until needed for the scanning/imaging process.

Customized scanning programs for all scannables were prepared to selectively read the student response booklets and to format the scanned information electronically according to predetermined requirements. Any information that had been designated time-critical or process-critical was handled first.

In addition to numerous real-time quality control checks, duplex read, a transport printer that prints a unique identifying number on each sheet of each booklet, and on-line editing capability, the 5000i scanners offer features that make them compatible with Internet technology.

SCANNING QUALITY CONTROL

NCS scanners are equipped with many built-in safeguards that prevent data errors. The scanning hardware is continually monitored for conditions that will cause the machine to shut down if standards are not met. It will display an error message and prevent further scanning until the condition is corrected. The areas monitored include document page and integrity checks, user-designed on-line edits, and many internal checks of electronic functions.

Before every scanning shift begins, Measured Progress operators perform a daily diagnostic routine. This is yet another step to protect data integrity and one that has been done faithfully for the many years that we have been involved in production scanning. In the rare event that the routine detects a photocell that appears to be out of range, we calibrate that machine and perform the test again. If the read is still not up to standard, we call for assistance from our field service engineer.

As a final safeguard, spot checks of scanned files, bubble by bubble and image by image, were routinely made throughout scanning runs. The result of these precautions, from the original layout of the scanning form to the daily vigilance of our operators, was a scan error rate well below 1 per 1000.

ELECTRONIC DATA FILES

Once the data had been entered and the scanning logs and other paperwork completed, the booklets themselves were put into storage (where they stayed for at least 180 days beyond the close of the fiscal year). When it had been determined that the files were complete and accurate, those files were duplicated electronically and made available for many other processing options.

CHAPTER 9—ITEM ANALYSES

As noted in Brown (1983), “a test is only as good as the items it contains.” A complete evaluation of a test’s quality must include an evaluation of each question. Both the *Standards for Educational and Psychological Testing* and the *Code of Fair Testing Practices in Education* include standards for identifying quality questions. Questions should assess only knowledge or skills that are identified as part of the domain being tested and should avoid assessing irrelevant factors. They should also be unambiguous and free of grammatical errors, potentially insensitive content or language, and other confounding characteristics. Further, questions must not unfairly disadvantage test takers from particular racial, ethnic, or gender groups.

Both qualitative and quantitative analyses are conducted to ensure that Montana CRT-Alternate items meet these standards. Qualitative analyses are described in earlier sections of this report; this section focuses on the more quantitative evaluations. The statistical evaluations included are: difficulty indices, item-test correlations, and differential item functioning (DIF) analyses. The item analyses presented here are based on the statewide administration of the Montana CRT-Alternate in spring 2005. About 114 grade 4 students, 127 grade 8 students, and 108 grade 10 students participated in the assessment.

DIFFICULTY INDICES (*P*)

All tasks were evaluated in terms of item difficulty according to standard classical test theory practices. Difficulty was defined as the average proportion of points achieved on an item, and was measured by obtaining the average score on an item and dividing by the maximum score for the item. Tasks are scored polytomously, where a student can achieve a score of 0, 1, 2, 3, or 4 for the item. By computing the difficulty index as the average proportion of points achieved, the items are placed on a scale that ranges from 0.0 to 1.0. Although this index is traditionally described as a measure of difficulty, it is properly interpreted as an “easiness index” because larger values indicate easier questions.

An index of 0.0 indicates that all students received no credit for the item, and an index of 1.0 indicates that all students received full credit for the item. Items that have either a very high or very low difficulty index are considered to be potentially problematic because they are either so difficult that few students get them right or so easy that nearly all students get them right. In either case, such items should be reviewed for appropriateness for inclusion on the assessment. If an assessment were comprised entirely of very easy or very hard items, all students would receive nearly the same scores and the assessment would not be able to differentiate high ability students from low ability students.

ITEM-TEST CORRELATIONS (ITEM DISCRIMINATION)

A desirable feature of an item is that the higher ability students perform better on the item than lower ability students. The correlation between student performance on a single item and total test score is a commonly used measure of this characteristic of an item. Within classical test theory, the item-test correlation is referred to as the item's discrimination because it indicates the extent to which successful performance on an item discriminates between high and low scores on the test. The discrimination index used to evaluate Montana CRT-Alternate tasks was the Pearson product-moment correlation. The theoretical range of this statistic is -1 to $+1$.

Discrimination indices can be thought of as measures of how closely a question assesses the same knowledge and skills assessed by other questions contributing to the criterion total score. That is, the discrimination index can be thought of as a measure of construct consistency. In light of this interpretation, the selection of an appropriate criterion total score is crucial to the interpretation of the discrimination index. For the Montana CRT-Alternate, the test total score was used as the criterion score.

SUMMARY OF ITEM ANALYSIS RESULTS

A summary of the item difficulty and item discrimination statistics for each grade/content combination is presented in Table 1 on the next page. The mean difficulty values shown in Table 1 indicate that, overall, students performed well on the items on the Montana CRT-

Alternate. Difficulty values for assessments designed for the general population (i.e., regular, rather than alternate, assessments) tend to be in the 0.4 to 0.7 range for the majority of items. Comparing these values to the difficulty values in Table 1 shows that students' performance on the CRT-Alternate items is clearly higher than is generally seen with regular assessments. Because the nature and purpose of alternate assessments are different from those of regular assessments, it is difficult to interpret this difference.

A similar pattern is visible with the mean discrimination values. Again, the values shown in the table are higher than would generally be seen with regular assessments. Part of the reason for this difference is the fact that all items on the CRT-Alternate are polytomously scored. In general, polytomous items will tend to have higher discrimination values than dichotomous (e.g., multiple-choice) items.

Because the nature and use of this assessment are different than those for general assessments, and because very few guidelines exist as to criteria for interpreting these values for alternate assessments, the statistics presented in Table 1 should be interpreted with caution.

Table 1
Item Analysis

Grade	Content Area	Difficulty		Discrimination	
		Mean	StDev	Mean	StDev
4	Reading	0.85	0.11	0.66	0.12
	Mathematics	0.78	0.12	0.65	0.15
8	Reading	0.88	0.08	0.70	0.12
	Mathematics	0.79	0.10	0.73	0.12
10	Reading	0.91	0.07	0.64	0.12
	Mathematics	0.84	0.10	0.67	0.19

DIFFERENTIAL ITEM FUNCTIONING

Investigations of item or test bias seek to determine whether scores for subgroups of students may be affected by attributes other than those the test is intended to measure. Such investigations usually begin by examining whether subgroups of students perform differently than expected on individual items. Specifically, differences due to irrelevant factors are examined. If such differential item functioning (DIF) is detected, a qualitative assessment of the item is made to determine whether the item is biased against a particular group. It should be noted that the detection of DIF does not imply that the item is biased; instead, it is a statistical tool that helps identify items that may be biased.

The *Code of Fair Testing Practices in Education* (1988) explicitly states that subgroup differences in performance due to irrelevant factors should be examined when sample size permits, and actions should be taken to make certain that differences in performance are due to construct-relevant, rather than irrelevant, factors. The *Standards for Educational and Psychological Testing* (1999) includes similar guidelines.

DIF procedures are designed to identify questions for which subgroups of interest perform differentially beyond the impact of differences in overall achievement. However, due to very small sample sizes (i.e., around 100 total students) it is unreasonable to calculate DIF statistics for the Montana CRT-Alternate. That is, Type I error rates would be unreasonably high and would result in incorrect conclusions regarding the functioning of the items between reference and focal groups. Thus, DIF statistics are not included as part of this technical report.

OPI was responsible for organizing and facilitating committees to review items and reading passages for bias and sensitivity. OPI sent the feedback from the committees to Measured Progress to make the appropriate changes to the items and reading passages.

CHAPTER 10—RELIABILITY

Although an individual question's performance is an important focus for evaluation, a complete evaluation of an assessment must also address the way questions function together and complement one another. Tests that function well provide an accurate assessment of the student's level of ability. Unfortunately, no test can do this perfectly. A variety of factors can contribute to a given student's score being either higher or lower than his or her true ability. Collectively, these extraneous factors that impact a student's score are referred to as measurement error. Any assessment includes some amount of measurement error; that is, no measurement can be perfectly accurate. This is true of academic assessments—no assessment can measure students perfectly accurately; some students will receive scores that underestimate their true ability, and other students will receive scores that overestimate their true ability. When tests have a high amount of measurement error student scores are very unstable. Students with high ability may get low scores or vice versa. Consequently, one cannot reliably tell a student's true level of ability with such a test. Assessments that have less measurement error (i.e., errors made are small on average and student scores on such a test will consistently represent their ability) are described as reliable.

There are a number of ways to estimate an assessment's reliability. One possible approach is to give the same test to the same students at two different points in time. If students receive the same scores on each test, then the extraneous factors affecting performance are small and the test is reliable (this is referred to as test-retest reliability). A potential problem with this approach is that students may remember questions from the first administration or may have gained (or lost) knowledge or skills in the interim between the two administrations. A solution to the 'remembering questions' problem is to give a different, but parallel test at the second administration. If student scores on each test correlate highly the test is considered reliable (this is known as alternate forms reliability, because an alternate form of the test is used in each administration). This approach, however, does not address the problem that students may have gained (or lost) knowledge or skills in the interim between the two administrations. In addition, the practical challenges of developing and administering parallel forms generally preclude the use of parallel forms reliability indices. One way to address these problems is to split the test in half and then correlate students' scores on the two half-

tests; this in effect treats each half-test as a complete test. By doing this, the problems associated with an intervening time interval, and of creating and administering two parallel forms of the test, are alleviated. This is known as a split-half estimate of reliability. If the two half-test scores correlate highly, questions on the two half-tests must be measuring very similar knowledge or skills. This is evidence that the questions complement one another and function well as a group. This also suggests that measurement error will be minimal.

The split-half method requires a judgment regarding the selection of which questions contribute to which half-test score. This decision may have an impact on the resulting correlation; different splits will give different estimates of reliability. Cronbach (1951) provided a statistic, α , that avoids this concern about the split-half method. Cronbach's α gives an estimate of the average of all possible splits for a given test. Cronbach's α is often referred to as a measure of internal consistency because it provides a measure of how well all the items in the test measure one single underlying ability. Cronbach's α is computed using the following formula:

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n s^2(Y_i)}{s_x^2} \right]$$

where i indexes the item
 n is the total number of items,
 $s^2(Y_i)$ represents individual item variance, and
 s_x^2 represents the total test variance

RELIABILITY

Table 2 on the next page presents Cronbach's α coefficient for each subject area (reading and mathematics), for each grade level. The values in Table 2 are all greater than 0.90, indicating that these tests have a high level of reliability. Note, however, that because the CRT Alternate is individually administered, the reliability values are likely to be inflated due to administrator effects.

Table 2

Reliability Analysis – All Grades

Grade	Content Area	Reliability
4	Mathematics	0.93
	Reading	0.91
8	Mathematics	0.95
	Reading	0.91
10	Mathematics	0.95
	Reading	0.92

CHAPTER 11—SCALING

TRANSLATING RAW SCORES TO SCALED SCORES AND PERFORMANCE LEVELS

Montana CRT-Alternate scores in each content area are reported on a scale that ranges from 200 to 300. Scaled scores supplement the Montana CRT-Alternate performance-level results by providing information about the position of a student's results within a performance level. School- and district-level scaled scores are calculated by computing the average of student-level scaled scores. Students' raw scores, or total number of points, on the Montana CRT-Alternate tests are translated to scaled scores using a data analysis process called **scaling**. Scaling simply converts raw points from one scale to another. In the same way that the same temperature can be expressed on either the Fahrenheit or Celsius scales and the same distance can be expressed either in miles or kilometers, student scores on the Montana CRT-Alternate tests could be expressed as raw scores (i.e., number right) or scaled scores.

It is important to note that converting from raw scores to scaled scores does not change the students' performance-level classifications. Given the relative simplicity of raw scores, it is fair to ask why scaled scores are used in Montana CRT-Alternate reports instead of raw scores. Foremost, scaled scores offer the advantage of simplifying the reporting of results across content areas, grade levels, and subsequent years. Because the standard-setting process typically results in different cut scores across content areas on a raw score basis, it is useful to transform these raw cut scores to a scale that is more easily interpretable and consistent. For the Montana CRT-Alternate, a score of 225 is the cut score between the **Novice** and **Nearing Proficiency** performance levels. This is true regardless of which content area, grade, or year one may be concerned with. If one were to use raw scores, the raw cut score between **Novice** and **Nearing Proficiency** may be, for example, 35 in mathematics at grade 8, but may be 33 in mathematics at grade 10, or 36 in reading at grade 8. Using scaled scores greatly simplifies the task of understanding how a student performed.

Cut points for the Montana CRT-Alternate were originally set at the standard setting held in July, 2004. (See Appendix C for the standard setting report for details on the standard

setting meeting.) The original cut scores were established on the raw score metric. However, some modifications were made to the assessments for the 2005 administration, so before calculating scaling coefficients for the 2005 tests, it was first necessary to find 2005 equivalents to the original cut points. The cut points were determined by matching the 2004 percents-in-category as closely as possible. Note that this procedure implies an assumption that student performance did not change from 2004 to 2005. Given the small numbers of students who took the CRT-Alternate, it was not possible to evaluate how tenable that assumption was. Therefore, no inferences can be made about changes in student preparedness between 2004 and 2005.

Once the 2005 raw score cut points were determined, the next step was to calculate the transformation coefficients that would be used to place students' raw scores onto the score scale used for reporting. As previously stated, student scores on the Montana CRT-Alternate are reported in integer values from 200 to 300 with three scores representing cut scores on each assessment. Two of the three cut points (novice/nearing proficiency and nearing proficiency/proficient) were pre-set at 225 and 250, respectively; the third cut point, between proficient and advanced, was allowed to vary across tests, depending on where the raw score cuts were placed. Allowing the upper cut to float results in a single conversion equation for each test, which simplifies interpretation of scaled scores and their summary statistics. Table 3 on the next page presents the scaled score range for each performance level in each grade/content area combination. Note that, for Grade 4 Mathematics, matching the 2005 percents-in-category to those observed in 2004 resulted in the upper cut being placed at 300. This is because a larger percentage of students score at the top of the raw score range on that test in 2005 than in 2004. However, as mentioned above, this change should not be over-interpreted, given the small number of students tested.

Table 3

Grade	Content Area	Scaled Score Range for each Performance Level			
		Novice	Nearing proficiency	Proficient	Advanced
4	Reading	200–224	225–249	250–257	258–300
	Mathematics	200–224	225–249	250–299	300
8	Reading	200–224	225–249	250–256	257–300
	Mathematics	200–224	225–249	250–271	272–300
10	Reading	200–224	225–249	250–264	265–300
	Mathematics	200–224	225–249	250–282	283–300

The scaled scores are obtained by a simple linear transformation of the raw scores using the values of 225 and 250 on the scaled score metric and the associated 2005 raw score cut points to define the transformation. The scaling coefficients were calculated using the following formulae:

$$b = 225 - m(x_1)$$

$$m = \frac{225 - 250}{x_1 - x_2}$$

where m is the slope of the line providing the relationship between the raw and scaled scores, b is the intercept, x_1 is the cut score on the raw score metric for the novice/nearing proficiency cut, and x_2 is the cut score on the raw score metric for the nearing proficiency/proficient cut. Scaled scores were then calculated using the following linear transformation:

$$ss = m(x) + b$$

where x represents a student's raw score. The values obtained using this formula were rounded to the nearest integer and truncated, as necessary, such that no student received a score below 200 or higher than 300.

CHAPTER 12—REPORTING

The CRT-Alternate assessments were designed to measure student performance against Montana's Content Standards and Expanded Benchmarks. Consistent with this purpose, results on the CRT-Alternate were reported in terms of performance levels that describe student performance in relation to these established state standards. There are four performance levels: *Advanced*, *Proficient*, *Nearing Proficiency*, and *Novice* (CRT-Alternate Performance Level Descriptors, Scaled Score Ranges, and Raw Scores are described in greater detail in Appendix "D"). Students receive a separate performance-level classification (based on total scaled score) in each content area.

School- and system-level results are reported as the number and percentage of students attaining each performance level at each grade level tested. Disaggregations of students are also reported at the school and system levels. The CRT-Alternate reports are

- Student Reports;
- Class Roster & Item-Level Reports;
- School Summary Reports; and
- System Summary Reports.

"Decision Rules" were formulated in late spring 2005 by OPI and Measured Progress to identify students, during the reporting process, to be excluded from school and system-level reports. A copy of these "Decision Rules" is included in this report as Appendix G.

State summary results were provided to OPI on confidential CDs and via a secure Web site. The report formats are included in Appendix F. These reports were shipped to System Test Coordinators on or before June 3, 2005 for distribution to schools within their respective systems/districts. System Test Coordinators and teachers were also provided with copies of the *Guide to Interpreting the 2005 Criterion-Referenced Test and CRT-Alternate Assessment Reports*, to assist them in understanding the connection between the assessment and the classroom. The guide provides information about the assessment and the use of assessment results.

CHAPTER 13—VALIDITY SUMMARY

The purpose of this manual is to describe several technical aspects of the CRT-Alternate in an effort to contribute to the accumulation of validity evidence to support CRT-Alternate score interpretations. Because it is the interpretations of test scores that are evaluated for validity, not the test itself, this manual presents documentation to substantiate intended interpretations (AERA, 1999). Each of the chapters in this manual contributes important information to the validity argument by addressing one or more of the following aspects of the CRT-Alternate: test development, test alignment, test administration, scoring, item analyses, reliability, scaling, performance levels and reporting.

The CRT-Alternate assessments are based on, and aligned to, Montana's Content Standards and Expanded Benchmarks in Reading and Mathematics. Intended inferences from the CRT-Alternate results are about student achievement on Montana's reading and mathematics content standards and expanded benchmarks, and these achievement inferences are meant to be useful for program and instructional improvement and as a component of school accountability.

The *Standards for Educational and Psychological Testing* (1999) provides a framework for describing sources of evidence that should be considered when constructing a validity argument. These sources include evidence based on the following five general areas: test content, response processes, internal structure, relationship to other variables, and consequences of testing. Although each of these sources may speak to a different *aspect* of validity, they are not distinct *types* of validity. Instead, each contributes to a body of evidence about the comprehensive validity of score interpretations.

A measure of test content validity is to determine how well the assessment tasks represent the curriculum and standards for each subject and grade level. This is informed by the item development process, including how the test blueprints and test items align to the curriculum and standards. Viewed through this lens provided by the Standards, evidence based on test content was extensively described in Chapters 2 through 7. Item alignment with Montana content standards; item bias, sensitivity and content appropriateness review processes;

adherence to the test blueprint; use of standardized administration procedures; and appropriate test administration training are all components of validity evidence based on test content. As discussed earlier, all CRT-Alternate test questions are aligned by Montana educators to specific Montana Content Standards, and undergo several rounds of review for content fidelity and appropriateness. Finally, tests are administered according to state-mandated standardized procedures, and all test administrators are required to review the training CD.

The scoring information in Chapter 8 describes the steps taken to train the teachers administering the assessment on scoring procedures, as well as quality control procedures related to scanning. In order to obtain additional validity evidence, it would be helpful to conduct a study in which a percentage of teachers administering the assessment would be videotaped to confirm validity of administration and scoring.

Evidence based on internal structure is presented in great detail in the discussions of item analyses and reliability in Chapters 9 and 10. Technical characteristics of the internal structure of the assessments are presented in terms of classical item statistics (item difficulty, item-test correlation) and reliability coefficients. In general, item difficulty and discrimination indices were in acceptable and expected ranges. Very few items were answered correctly at near-chance or near-perfect rates. Similarly, the positive discrimination indices indicate that most items were assessing consistent constructs, and students who performed well on individual items tended to perform well overall.

Evidence based on the consequences of testing is addressed in the scaling and reporting information in Chapters 11 and 12, as well as in the test interpretation guide, which is a separate document that is referenced in the discussion of reporting. Each of these chapters speaks to the efforts undertaken to promote accurate and clear information provided to the public regarding test scores. Scaled scores offer the advantage of simplifying the reporting of results across content areas, grade levels, and subsequent years. Performance levels provide users with reference points for mastery at each grade level, which is another useful and simple way to interpret scores. Several different standard reports are provided to stakeholders. Additional evidence of the consequences of testing could be supplemented with broader investigation of the impact of testing on student learning.

To further support the validity argument, additional studies to provide evidence regarding the relationship of CRT-Alternate results to other variables include the extent to which scores from the CRT-Alternate assessments converge with other measures of similar constructs, and the extent to which they diverge from measures of different constructs. Relationships among measures of the same or similar constructs can sharpen the meaning of scores and appropriate interpretations by refining the definition of the construct.

The evidence presented in this manual supports inferences of student achievement on the content represented on the Montana Content Standards for Reading and Mathematics for the purposes of program and instructional improvement and as a component of school accountability.

SECTION IV—REFERENCES

American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd ed.). Fort Worth, TX: Holt, Rinehart, and Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.

Joint Committee on Testing Practices (1988). *Code of fair testing practices in education*. Washington, DC: National Council on Measurement in Education.

SECTION V: APPENDICES

- A. Advisory Committee Members, Test Development and Revision Workshop Participants, and Beta Test Participants**
- B. Technical Advisory Committee**
- C. Standard Setting Report, Evaluation Summaries, and PowerPoint Presentations**
- D. CRT Performance Level Descriptors, Scaled Scores, and Raw Scores**
- E. CRT-Alternate Released Performance Indicators**
- F. Report Shells**
- G. Reporting Decision Rules**

Appendix A

CRT-ALTERNATE ASSESSMENT ADVISORY COMMITTEE

Kim Allen, Great Falls, MT
Nancy Anderson, Great Falls, MT
Susan Gregory, Billings, MT
Joanne Hallock, Fort Peck, MT

Shaun Harrington, Billings, MT
Carol Kron, Livingston, MT
Joyce Silverthorne, Dixon, MT
Karla Wohlwend, Havre, MT

CRT-ALTERNATE ASSESSMENT MATH DEVELOPERS

Jenny Bland, Libby, MT
Lee Brown, Missoula, MT
Susan Buechler, Billings, MT
Janet Euell, Ballantine, MT
Dr. Roberta J. Flexer, Louisville, CO
Rebecca Frisbee, Great Falls, MT
Karen Johnson, Deer Lodge, MT
Kelly Klein, Worden, MT

Carol Kron, Livingston, MT
Judith McKay, Libby, MT
Denielle Miller, Bozeman, MT
Joyce Miller, Great Falls, MT
Karen Nave, Havre, MT
Mary Nelson, Livingston, MT
Diane Sherman, Huntley, MT
Glenda Truesdell, Townsend, MT

CRT-ALTERNATE ASSESSMENT READING DEVELOPERS

Jerri Boksich, Whitefish, MT
Teri Brogdon, Denver, CO
Sue Brown, Whitefish, MT
Glenn Castle, Cut Bank, MT
Sandy Grey Eagle, Wibaux, MT
Norma MacKenzie, Whitefish, MT
Linda Malingo, Kalispell, MT

Terrie Noser, Libby, MT
Michele Paine, Bigfork, MT
Juanita Sloss, Browning, MT
Debra Waite, Bozeman, MT
Kristen Walser, Bozeman, MT
Robin Zeal, Whitefish, MT

CRT-ALTERNATE ASSESSMENT REVISION WORKSHOP PARTICIPANTS

Theresa Anderson, Billings, MT
Jenny Bland, Libby, MT
Sandy Grey Eagle, Wibaux, MT
Darlene Kolczak, Zortman, MT
Carol Kron, Livingston, MT
Dan Laughlin, Anaconda, MT

Judy McKay, Libby, MT
Karen Nave, Havre, MT
Sheila Ryan, Lame Deer, MT
Marjorie Stricklin, Great Falls, MT
Holly Wick, Great Falls, MT

CRT-ALTERNATE BETA TEST TEACHERS AND ADMINISTRATORS

Emilie Anderson, Missoula, MT
Tara Bohn, Missoula, MT
Tammy Cole, Lolo, MT
Marie Craton, Missoula, MT
Maureen Dachs, Kalispell, MT
Geri Darko, Great Falls, MT
Anne Fitz, Helena, MT
Heidi Foreman, Helena, MT
Jeanne Glendening, Missoula, MT
Stacey Hanson, Polson, MT
Wendy Ihde, Frenchtown, MT

Susan Jacobson, Kalispell, MT
Bill Johnson, Columbia Falls, MT
Lisa Lowney, Helena, MT
Bonnie McCormick, Stevensville, MT
Bev McDaniels, Hamilton, MT
Janet Mullis, Kalispell, MT
WyAnn Northrop, Missoula, MT
Paula Onstad, Stevensville, MT
Megan Richert, Helena, MT
Criss Rigby, Philipsburg, MT
Marvin Williams, Helena, MT

Appendix B

Technical Advisory Committee (TAC) Members				
First Name	Last Name	Position	Department	Organization
Art	Bangert, Ph.D.	Assistant Professor	Adult and Higher Education	Montana State University
Liz	Burton, Ph.D.	Psychometrician	MDA	Measured Progress
Tim	Crockett	Vice President, Client Services		Measured Progress
Carolyn	Haug, Ph.D.	Asst. Division Director		Measured Progress
Michael	Kozlow, Ph.D.	Director		Northwest Regional Ed. Lab
Scott	Marion, Ph.D.	Consultant		Center for Assessment
Mike	Nering, Ph.D.	Psychometrician	MDA	Measured Progress
Madalyn	Quinlan	Chief Executive Officer		OPI
Stanley	Rabinowitz, Ph.D.	Program Director	Assessment & Standards Development Services	WestEd
Nam	Raju, Ph. D.	Distinguished Professor		Institute of Psychology
Steve	Sireci, Ph.D.	Associate Professor		UMASS Amherst
Judy	Snow	State Assessment Director		OPI
Wes	Snyder, Ph.D.	Assistant Vice Pres.	Research & Director of Office of International Programs	University of Montana
Kevin	Sweeney, Ph.D.	Division Director	MDA	Measured Progress
Rebecca	Walk	Division Director	Special Education	Measured Progress
Bud	Williams	Asst. Superintendent		OPI

Appendix C

Montana CRT-Alternate Standard-Setting Report June 30 & July 1, 2004 Helena, MT

OVERVIEW OF STANDARD-SETTING MEETING

The standard-setting meeting held to establish cut scores in reading and mathematics, grades 4, 8 and 10, on the Montana CRT-Alternate was held Wednesday, June 30, and Thursday, July 1. There were six panels of seven to ten panelists each, and each panel completed the standard-setting activities over the course of the two days.

The standard-setting method implemented for both content areas and all grades was a modified version of the body-of-work method. An overview of this method is described below. All panels followed the same procedures.

To help ensure consistency of procedures between panels, each panel was led through the standard setting process by trained facilitators from Measured Progress.

OVERVIEW OF PROCESS

This section of the report provides an overview of the standard-setting process as it was implemented in Montana. The process was divided into the following three stages, each with several constituent tasks:

❖ Tasks completed prior to the meeting

- Creation of performance levels and performance level definitions
- Preparation of materials for panelists
- Preparation of presentation materials
- Preparation of systems and materials for analysis during the meeting
- Selection of panelists

❖ Tasks completed during the meeting

- Orientation
- Review of assessment materials
- Review of performance level definitions
- Round 1 judgments
- Tabulation of round 1 results
- Round 2: Comparison of panelist judgments and opportunity for revised judgments
- Tabulation of round 2 results
- Round 3: Comparison of panelist results and impact data, and final opportunity to revise judgments
- Modification of performance level definitions
- Evaluation

❖ Tasks completed after the meeting

- Analysis and review of panelists' feedback
- Preparation of recommended cut scores
- Preparation of standard-setting report

TASKS COMPLETED PRIOR TO THE STANDARD-SETTING MEETING

CREATION OF PERFORMANCE LEVELS AND PERFORMANCE LEVEL DEFINITIONS

The performance level definitions provided panelists the official description of the knowledge, skills and abilities students are expected to be able to display to be classified into each performance level. These performance level definitions were presented to panelists. The definitions are provided in Appendix A of this document.

PREPARATION OF MATERIALS FOR PANELISTS

The following materials were assembled into folders for presentation to the panelists at the standard-setting meeting:

- Meeting agenda
- Confidentiality agreement
- Performance level definitions
- Assessment protocol
- Scoring rubrics
- Visual item map
- Student profiles
- Rating forms
- Evaluation form

PREPARATION OF PRESENTATION MATERIALS

The PowerPoint presentations used in the opening session was prepared prior to the meeting.

PREPARATION OF SYSTEMS AND MATERIALS FOR ANALYSIS DURING THE MEETING

The programming of all analyses to be conducted during the standard-setting meeting was completed and thoroughly tested prior to the standard-setting meeting.

SELECTION OF PANELISTS

Panelists were selected prior to the standard-setting meeting. The goal was to have 10 panelists for each of the six panels, for a total of 60. The actual number of panelists who participated was 53, distributed as follows:

- Grade 4 Math – 9
- Grade 8 Math – 8
- Grade 10 Math – 9
- Grade 4 Reading – 10
- Grade 8 Reading – 7
- Grade 10 Reading – 10

Of the 53 panelists, there were 44 teachers, 4 administrators, and 5 other (parents, librarians, counselors, etc.) All panelists were white, and 49 were female and 4 male. Some CRT-Alternate panelists were drawn from among those CRT applicants who had special education experience.

TASKS COMPLETED DURING THE STANDARD-SETTING MEETING

ORIENTATION

The standard-setting meeting on day 1 began with a general orientation session that was attended by all panelists. The purpose of this session was to provide some background information, provide an introduction to the issues of standard setting, to explain the activities that would occur during the standard-setting meeting, and to go over some of the materials that would be used. In addition, some video clips were shown to familiarize the panelists with the assessment and scoring processes. At the conclusion of the opening session the floor was opened to questions about the standard-setting process. Some of the questions focused on the uses of the standard-setting results and other policy related issues, and some questions were about the ratings to be made and clarification of the process.

After the large-group session, the panelists assembled into grade/content area groups. Each group was in a separate room and each room was further divided into two tables of three to five panelists each.

REVIEW ASSESSMENT MATERIALS

Once in the smaller groups, the panelists carefully reviewed the assessment protocol, scoring rubric and scaffolding directions for each indicator.

REVIEW PERFORMANCE LEVEL DEFINITIONS

In the next step, panelists reviewed the performance level definitions. They then discussed the specific characteristics that students in each performance level would have. Specifically, they determined the knowledge, skills and abilities that students would need to demonstrate in order to be classified into each performance level category. Once the panelists came to consensus, the descriptors were written onto chart paper and posted in the room so the panelists could refer to them throughout the standard-setting process.

ROUND 1 JUDGMENTS

In the first round, subject area panelists worked individually to make their initial ratings. Each panelist was given a set of student profiles that were prepared by Measured Progress prior to the standard setting. To create the profiles, MP staff first divided the full range of raw score points into a number of score intervals; the number of intervals for the different tests ranged from 16 to 25. Second, the average item score of all students whose total score fell within each interval was calculated for each item. Thus, each student profile showed the expected performance on each item of a student whose total raw score fell in the middle of the score interval on which that profile was based.

In doing the round 1 ratings, panelists used the performance level definitions with the expanded information they completed in the previous step as well as the student profiles. In addition, panelists could refer to the visual item map. The visual item map gave a visual representation of student performance on each of the indicators and could help panelists understand the relationship among the different indicators. Starting with the first (lowest scoring) profile, each panelist considered that student's pattern of responses and what it indicated about the knowledge, skills and abilities that student had, then decided which

performance level that student should be classified into. They then repeated the process for each profile in turn. Each panelist used the rating form provided to record his/her ratings.

TABULATION OF ROUND 1 RESULTS

Each table of panelists received a feedback form that showed how each panelist at the table rated each of the profiles, as well as the average cut points for the table. The average cut points were determined using logistic regression. Specifically, for a given cut, each panelist's rating for each student profile was dichotomized (i.e., above or below the cut). A logistic function was fit to the data for that cut, and the point of inflection on this curve was used to establish the average cut point on the raw score scale. The process was then repeated for the remaining cuts. This information was used to facilitate discussion of the table ratings in round 2.

ROUND 2: COMPARISON OF PANELIST JUDGMENTS AND OPPORTUNITY FOR REVISED JUDGMENTS

During round 2, the panelists at each table examined the results from round 1 and discussed their ratings. The panelists shared the rationale for their ratings in terms of the knowledge and skills students must demonstrate in order to be categorized into a particular performance level. After all panelists had an opportunity to discuss their ratings and each table of panelists completed its discussions, the panelists then had the opportunity to change or revise their round 1 ratings. Each panelist once again used the rating form to record his/her ratings.

TABULATION OF ROUND 2 RESULTS

As with round 1, a feedback form was provided to each table after round 2 showing the ratings of each panelist and the average cut scores for the table. In addition, for round 2, the average cut scores for the room as a whole were also provided, along with impact data showing the percentage of students that would be placed into each performance level if the room average cut scores were used.

ROUND 3: COMPARISON OF PANELIST RESULTS AND IMPACT DATA, AND FINAL OPPORTUNITY TO REVISE JUDGMENTS

All of the tabulated results from round 2 were distributed to panelists prior to the final round of ratings. As a whole room, panelists discussed the round 2 ratings and the impact data. After the round 3 discussions, each panelist had another opportunity to change or revise his/her ratings, using the rating form.

The cut scores and percentage of students classified into each performance level, based on the group average cut score from round 3, are presented in Tables 1 through 6.

Table 1
Cut Scores and Impact Data – Grade 4 Reading

Proficiency Level	Minimum Score	% in Level
Advanced	81	38
Proficient	58	38
Nearing Proficiency	26	20
Novice	--	5

Table 2
Cut Scores and Impact Data – Grade 8 Reading

Proficiency Level	Minimum Score	% in Level
Advanced	96	31
Proficient	78	38
Nearing Proficiency	42	19
Novice	--	12

Table 3

Cut Scores and Impact Data – Grade 10 Reading

Proficiency Level	Minimum Score	% in Level
Advanced	99	37
Proficient	75	38
Nearing Proficiency	39	12
Novice	--	13

Table 4

Cut Scores and Impact Data – Grade 4 Math

Proficiency Level	Minimum Score	% in Level
Advanced	108	13
Proficient	74	55
Nearing Proficiency	44	10
Novice	--	22

Table 5

Cut Scores and Impact Data – Grade 8 Math

Proficiency Level	Minimum Score	% in Level
Advanced	115	26
Proficient	85	31
Nearing Proficiency	46	15
Novice	--	28

Table 6
Cut Scores and Impact Data – Grade 10 Math

Proficiency Level	Minimum Score	% in Level
Advanced	118	38
Proficient	77	29
Nearing Proficiency	37	14
Novice	--	18

MODIFICATION OF PERFORMANCE LEVEL DEFINITIONS

After completing the rating process, the panelists listed suggested modifications to the performance level definitions based on the round 3 results of the standard-setting process.

EVALUATION

At the end of the process, panelists anonymously completed an evaluation form. The results of the evaluation are presented on pages 63-81.

TASKS COMPLETED AFTER THE STANDARD-SETTING MEETING

Upon conclusion of the standard-setting meeting, several important tasks were completed. These tasks centered on reviewing the standard-setting meeting and addressing anomalies that may have occurred in the process or in the results.

ANALYSIS AND REVIEW OF PANELISTS' FEEDBACK

Upon completion of the evaluation forms, panelists' responses were reviewed. This review did not reveal any anomalies in the standard-setting process or indicate any reason that a particular panelist's data should not be incorporated in obtaining the final results. It appeared that all panelists understood the rating task and attended to it appropriately. Panelist responses to the evaluation items are presented on pages 63-81.

PREPARE RECOMMENDED CUT SCORES

The recommended cut scores coming out of the standard setting process are the results from round 3. These cut scores will be reviewed by the Montana Technical Advisory Committee and OPI for approval or modification.

PREPARATION OF STANDARD-SETTING REPORT

This report documents the procedures and results of the standard-setting meetings in the establishment of performance standards for the Montana CRT-Alternate.

TECHNICAL ADVISORY COMMITTEE (TAC)

Montana's Technical Advisory Committee (TAC) met on July 12 and 13, 2004, to review the standard-setting selection process, Standard-Setting Report, facilitator scripts, panelists' recommended cut scores, and panelists' evaluation summaries. A list of TAC members is included in this report as Appendix B.

TAC members reviewed panelists' judgments plus the standard error of measurement at confidence intervals. The TAC approved the standard-setting process, the modified body-of-work method, and procedures applied by Measured Progress and Montana educators. In addition, the TAC members provided OPI with cut score recommendations. OPI set the final cut scores in all grades and content areas (see Appendix D: CRT-Alternate Performance Level Descriptors and Cut Scores).

Standard Setting: Evaluation Summaries

READING GRADE 4

Evaluation of the Standard-Setting Procedures for the Montana Alternate Assessment

1. What is your overall impression of the process used to set performance standards for the Montana Alternate Assessment? (Circle one.)

- A. Very good (4)
- B. Good (5)
- C. Unsure (1)
- D. Poor (0)
- E. Very poor (0)

2. How clear were you with the performance level descriptors? (Circle one.)

- A. Very clear (7)
- B. Clear (2)
- C. Somewhat clear (1)
- D. Not clear (0)

3. How would you judge the length of time of this meeting for setting performance standards? (Circle one.)

- A. About right (10)
- B. Too little time (0)
- C. Too much time (0)

4. What factors influenced the standards you set? (For each, circle the most appropriate rating from 1=Not at All Influential to 5=Very Influential.)

- A. The performance level descriptors

Not at All Influential		Moderately Influential		Very Influential
1 (1)	2 (0)	3 (3)	4 (3)	5 (3)

- B. The assessment items (performance indicators)

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (0)	3 (3)	4 (2)	5 (5)

Other panelists

Not at All Influential		Moderately Influential		Very Influential
1 (1)	2 (1)	3 (4)	4 (3)	5 (1)

C. My experience in the field

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (0)	3 (0)	4 (4)	5 (6)

D. Other (please specify _____)

Not at All Influential		Moderately Influential		Very Influential
1 (0)				
2 (0)				
3 (0)				
4 (0)				
5 (0)				

5. Do you believe the cut scores set by the panel are correctly placed on the assessment score scale?

- A. Definitely yes (9)
- B. Probably yes (1)
- C. Unsure (0)
- D. Probably no (0)
- E. Definitely no (0)

Please explain your answer:

- *With the assessments that we were looking at, the cut scores fell into place*
- *Good argument and good discussion. Freedom to express ideas*
- *We were fairly close on the 1st cut, the cuts were a natural process*
- *With the combined scores used, they are set, could have had more scores to be more accurate*
- *The discussion and give and take*
- *After defining criteria, I feel our scores fit very well*
- *Work on the "why" part – Good mix of panelists*

6. How could the standard-setting process have been improved?

- *Perhaps more structure, but overall don't change*
- *Our group got along very well and we only needed 1.5 days. However, if there had been a lot of disagreements, we would have probably needed the full 2 days.*

For each statement below, please circle the rating that best represents your judgment.

7. The opening session was:

Not at All Useful				Very Useful
1 (0)	2 (1)	3 (2)	4 (6)	5 (1)

8. The performance level descriptors were:

Not at All Clear				Very Clear
1 (0)	2 (1)	3 (1)	4 (5)	5 (3)

9. Providing additional details to the performance level descriptors was:

Not at All Useful				Very Useful
1 (0)	2 (1)	3 (0)	4 (4)	5 (5)

10. The discussion with other panelists was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (1)	4 (0)	5 (9)

11. The student profile rating task was:

Not at All Clear				Very Clear
1 (0)	2 (0)	3 (1)	4 (5)	5 (4)

12. The impact data at the beginning of round 3 was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (1)	4 (3)	5 (5)
<i>One participant did not respond</i>				

Additional Comments

13. Please provide any additional comments or suggestions about the standard-setting process.

- *Thank you for the information and help with understanding the process*
- *Very interesting*
- *A wonderful process – room to express ideas and freedom to disagree*
- *Use these same people that have been involved so we can see the whole process*
- *People in our district (teachers)*

READING

GRADE 8

Evaluation of the Standard-Setting Procedures for the Montana Alternate Assessment

1. What is your overall impression of the process used to set performance standards for the Montana Alternate Assessment? (Circle one.)

- A. Very good (1)
- B. Good (6)
- C. Unsure (0)
- D. Poor (0)
- E. Very poor (0)

2. How clear were you with the performance level descriptors? (Circle one.)

- A. Very Clear (5)
- B. Clear (2)
- C. Somewhat Clear (0)
- D. Not Clear (0)

3. How would you judge the length of time of this meeting for setting performance standards? (Circle one.)

- A. About right (7)
- B. Too little time (0)
- C. Too much time (0)

4. What factors influenced the standards you set? (For each, circle the most appropriate rating from 1=Not at all Influential to 5=Very Influential.)

- A. The performance level descriptors

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (0)	3 (4)	4 (2)	5 (1)

- B. The assessment items (performance indicators)

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (2)	3 (4)	4 (1)	5 (0)

- C. Other panelists

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (2)	3 (3)	4 (2)	5 (0)

D. My experience in the field

Not at All Influential	Moderately Influential		Very Influential	
1 (0)	2 (1)	3 (1)	4 (3)	5 (2)

E. Other (please specify _____)

Not at All Influential	Moderately Influential		Very Influential	
1 (0)				
2 (0)				
3 (0)				
4 (0)				
5 (0)				

5. Do you believe the cut scores set by the panel are correctly placed on the assessment score scale?

- F. Definitely Yes (4)
- G. Probably Yes (3)
- H. Unsure (0)
- I. Probably No (0)
- J. Definitely No (0)

Please explain your answer:

- ☐ *I feel comfortable with these scores*
- ☐ *Yes because the group and individual process used to help attain this*
- ☐ *The scores provide success in testing, but not overly so*
- ☐ *A thorough evaluation of each profile was done and specific criteria for each level was determined*
- ☐ *Good discussions produce better understanding, pooling of ideas and information*
- ☐ *A lot of thought was put into the process. 95-100% of students are proficient and this is where they should be. More reliable data.*

6. How could the standard-setting process have been improved?

- ☐ *Longer time, more exemplars*
- ☐ *Larger number of panelists, more advanced notice may encourage participants*
- ☐ *I thought the process of setting standards was informative and flowed smoothly*
- ☐ *Explanation of the pink, purple and test booklet use together*
- ☐ *This was my first time to attend a standards setting process. I liked how the process is done and would like to implement processes like this in my job, whatever the topic may be!*

For each statement below, please circle the rating that best represents your judgment.

7. The opening session was:

Not at All Useful				Very Useful
1 (0)	2 (2)	3 (2)	4 (1)	5 (2)

8. The performance level descriptors were:

Not at All Clear				Very Clear
1 (0)	2 (0)	3 (4)	4 (3)	5 (0)

9. Providing additional details to the performance level descriptors was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (1)	4 (4)	5 (2)

10. The discussion with other panelists was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (0)	4 (4)	5 (3)

11. The student profile rating task was:

Not at All Clear				Very Clear
1 (0)	2 (0)	3 (1)	4 (7)	5 (1)

12. The impact data at the beginning of round 3 was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (2)	4 (2)	5 (3)

Additional Comments

13. Please provide any additional comments or suggestions about the standard-setting process.

- *I think the process was good. A lot of problems arrived because of the test questions themselves.*
- *The test needs serious revisions – restating test questions – questions that perhaps target specific skills that would lend itself to measuring growth*
- *It was an interesting process and enjoyable to participate in*
- *Important to involve more experienced teachers (larger # of panelists). Seems to be a general feeling among some participants that our work will count very minimally toward state standards decisions.*

READING GRADE 10

Evaluation of the Standard-Setting Procedures for the Montana Alternate Assessment

1. What is your overall impression of the process used to set performance standards for the Montana Alternate Assessment? (Circle one.)
 - A. Very good (4)
 - B. Good (4)
 - C. Unsure (1)
 - D. Poor (1) indicating at the opening session that our input “might” not be where standards are set – made up
 - E. Very poor (0)

2. How clear were you with the performance level descriptors? (Circle one.)
 - A. Very clear (2)
 - B. Clear (5)
 - C. Somewhat clear (3)
 - D. Not clear (0)

3. How would you judge the length of time of this meeting for setting performance standards? (Circle one.)
 - A. About right (9)
 - B. Too little time (0)
 - C. Too much time (1)

4. What factors influenced the standards you set? (For each, circle the most appropriate rating from 1=Not at all Influential to 5=Very Influential.)
 - A. The performance level descriptors

Not at All Influential	Moderately Influential		Very Influential	
1 (0)	2 (0)	3 (2)	4 (5)	5 (3)

 - B. The assessment items (performance indicators)

Not at All Influential	Moderately Influential		Very Influential	
1 (0)	2 (0)	3 (3)	4 (4)	5 (3)

C. Other panelists

Not at All Influential	Moderately Influential		Very Influential	
1 (0)	2 (1)	3 (5)	4 (2)	5 (2)

D. My experience in the field

Not at All Influential	Moderately Influential		Very Influential	
1 (0)	2 (3)	3 (0)	4 (3)	5 (4)

E. Other (please specify_____)

Not at All Influential	Moderately Influential		Very Influential	
1 (0)				
2 (0)				
3 (0)				
4 (0)				
5 (1)	Need for discussion			

5. Do you believe the cut scores set by the panel are correctly placed on the assessment score scale?

- K. Definitely yes (2)
- L. Probably yes (5)
- M. Unsure (2)
- N. Probably no (0)
- O. Definitely no (1)

Please explain your answer:

- *No, not at all – the panel administered the test*
- *I believe there is really a lot of room for “play” in the ratings as it depends so much on the teacher and the student*
- *I do not know the final results yet, but am comfortable with the discussion*
- *I would hope that they are what the standards are set at*
- *We had good conversations discussing the realities of the cut offs – I feel good about the changes we made*
- *There are many subjective aspects of the test. When future tests are assembled, give everything – except modifications for a particular student*
- *Yes, I believe these cut scores are correct because all have a great balance of educators!*
- *We agreed after discussions. The discussions were good.*
- *With the input of experienced test givers, we are definitely satisfied with the cuts*

6. How could the standard-setting process have been improved?

- *A cross analysis of test items to standards would be helpful. Most of us felt the items didn't do a very good job of getting to the standards, let alone attending to a school's curriculum*
- *The benchmarks didn't match with the questions.*
- *We think the test is not an accurate assessment of the benchmarks necessary. There are many questions that are a stretch!*
- *An interesting and/or professional experience*
- *A group consensus would have been helpful for my need for definite closure*
- *I don't have enough experience to tell you how to improve. I felt it went well.*
- *I had some problems with the expanded benchmarks – they were not sequential, i.e., predicting a much higher level skill*
- *We were a little confused at first – it needed to be made more clear at the beginning of the breakout that we were just to “explore” the test and descriptors*
- *Perhaps more professional guidance from the Measured Progress people, statisticians and OPI on the ramifications of this process to our students and teachers.*
- *Too many variables on test materials*
- *The psychometrician was too fast, too much. Rest was good.*

For each statement below, please circle the rating that best represents your judgment.

7. The opening session was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (1)	4 (6)	5 (2)
<i>Too long, redundant</i>				

8. The performance level descriptors were:

Not at All Clear				Very Clear
1 (0)	2 (0)	3 (3)	4 (5)	5 (2)

9. Providing additional details to the performance level descriptors was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (0)	4 (7)	5 (3)

10. The discussion with other panelists was:

Not at all Useful				Very Useful
1 (0)	2 (0)	3 (0)	4 (1)	5 (9)

11. The student profile rating task was:

Not at All Clear				Very Clear
1 (0)	2 (0)	3 (3)	4 (5)	5 (2)

12. The impact data at the beginning of round 3 was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (3)	4 (3)	5 (4)

Additional Comments

13. Please provide any additional comments or suggestions about the standard-setting process.
- *There appears there are many, many variables when giving the test. Materials used can possibly affect how an answer is scored. This affected perceptions when doing cuts.*
 - *I can understand the attempt to assess even our most severely handicapped students but to try and statistically analyze the group is “futile” as it then skews the group to reflect something that is NOT accurate because the students are so INDIVIDUAL.*
 - *Deanne was very helpful!*
 - *Thanks for letting me be a part of this team. It is difficult to make this a standardized process with the activities and variables being suggested us mandated.*
 - *With so many variables in suggested activities and in teacher attention/awareness, it is hard to find validity and reliability, i.e., standardized testing*
 - *I feel I learned a lot about this assessment. The test itself was unique however some changes need to be made. On the level descriptors, the words limited and moderate can be interpreted differently. Giving suggestions would help*
 - *Thank you!*
 - *The individual child defines his or her disability and we must move toward looking at the individual child and classifying and standardizing him based on his disability.*
 - *Very well organized and directions were clear. Long-range ramifications are still unclear to me. I will be anxious to see the results and see how it affects my district.*
 - *It felt like we were here to validate what Measured Progress had already done and would do. Our discussion leader was awesome!*

MATH GRADE 4

Evaluation of the Standard-Setting Procedures for the Montana Alternate Assessment

1. What is your overall impression of the process used to set performance standards for the Montana Alternate Assessment? (Circle one.)

- A. Very good (1)
- B. Good (4)
- C. Unsure (3)
- D. Poor (0)
- E. Very poor (0)

2. How clear were you with the performance level descriptors? (Circle one.)

- A. Very clear (0)
- B. Clear (6)
- C. Somewhat clear (1)
- D. Not clear (1)

3. How would you judge the length of time of this meeting for setting performance standards? (Circle one.)

- A. About right (5)
- B. Too little time (0)
- C. Too much time (3)

4. What factors influenced the standards you set? (For each, circle the most appropriate rating from 1=Not at all Influential to 5=Very Influential.)

- A. The performance level descriptors

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (2)	3 (1)	4 (4)	5 (1)

- B. The assessment items (performance indicators)

Not at All Influential		Moderately Influential		Very Influential
1 (1)	2 (0)	3 (1)	4 (5)	5 (1)

- C. Other panelists

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (0)	3 (4)	4 (2)	5 (2)

D. My experience in the field

Not at All Influential	Moderately Influential		Very Influential	
1 (1)	2 (0)	3 (1)	4 (2)	5 (4)

E. Other (*please specify*_____)

Not at All Influential	Moderately Influential		Very Influential	
1 (0)	2 (0)	3 (0)	4 (0)	5 (1) <i>Ethical Decisions</i>

5. Do you believe the cut scores set by the panel are correctly placed on the assessment score scale?

- A. Definitely yes (0)
- B. Probably yes (8)
- C. Unsure (0)
- D. Probably no (0)
- E. Definitely no (0)

Please explain your answer:

- *Seemed the right place*
- *I liked that we did work individually, as a table, than as a room. Good discussion.*
- *It makes as much sense as anything on a non-sensical test.*
- *Student's progress is really based on the level of assistance and the line of demarcation between the skills is not definitive enough.*

6. How could the standard-setting process have been improved?

- *The process was fairly straightforward to guide us to a pre-determined conclusion.*
- *The process seemed to be directed to a foregone conclusion. Is it possible to be more open-ended?*
- *One day is probably enough time. The process by tables the first day was more frustrating than helpful. Had the room worked as a whole, we could have gotten to the point that we did more expediently and then move on. It seemed like a great deal of wasted time that first day.*
- *I don't know – more people involved that are special educators or some knowledge of the test and how it is administered.*
- *Have more teachers who had given the test.*

For each statement below, please circle the rating that best represents your judgment.

7. The opening session was:

Not at All Useful				Very Useful
1 (0)	2 (1)	3 (3)	4 (3)	5 (1)

8.	The performance level descriptors were:				
	Not at All Clear				Very Clear
	1 (0)	2 (2)	3 (4)	4 (2)	5 (0)
9.	Providing additional details to the performance level descriptors was:				
	Not at All Useful				Very Useful
	1 (0)	2 (1)	3 (2)	4 (4)	5 (1)
10.	The discussion with other panelists was:				
	Not at All Useful				Very Useful
	1 (0)	2 (0)	3 (0)	4 (4)	5 (4)
11.	The student profile rating task was:				
	Not at all Clear				Very Clear
	1 (0)	2 (1)	3 (3)	4 (3)	5 (1)
12.	The impact data at the beginning of round 3 was:				
	Not at All Useful				Very Useful
	1 (0)	2 (0)	3 (1)	4 (6)	5 (1)

Additional Comments

13. Please provide any additional comments or suggestions about the standard-setting process.
- *Please continue to discuss the ethics of even giving this test and the process of reporting out the results to the public.*
 - *I realize that the assessment process is legally mandated, however, investing the money and time in this process seems wasted since these individuals will not be tax-paying persons.*
 - *Judy visited with our group and helped us work through our problems with testing this population of students. The reporting of this population with the regular population is misleading.*

MATH GRADE 8

Evaluation of the Standard-Setting Procedures for the Montana Alternate Assessment

1. What is your overall impression of the process used to set performance standards for the Montana Alternate Assessment? (Circle one.)

- A. Very good (3)
- B. Good (5)
- C. Unsure (0)
- D. Poor (0)
- E. Very poor (0)

2. How clear were you with the performance level descriptors? (Circle one.)

- A. Very clear (2)
- B. Clear (5)
- C. Somewhat clear (1)
- D. Not clear (0)

3. How would you judge the length of time of this meeting for setting performance standards? (Circle one.)

- A. About right (7)
- B. Too little time (0)
- C. Too much time (1)

4. What factors influenced the standards you set? (For each, circle the most appropriate rating from 1=Not at all Influential to 5=Very Influential.)

- A. The performance level descriptors

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (0)	3 (1)	4 (1)	5 (6)

- B. The assessment items (performance indicators)

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (0)	3 (1)	4 (3)	5 (4)

- C. Other panelists

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (0)	3 (5)	4 (0)	5 (3)

D. My experience in the field

Not at All Influential	Moderately Influential	Very Influential
1 (0)	2 (0) 3 (3) 4 (0)	5 (5)

E. Other (please specify_____)

Not at All Influential	Moderately Influential	Very Influential
1 (0)		
2 (0)		
3 (1)	<i>Mixture of group</i>	
4 (0)		
5 (2)	<i>Opportunities to make suggestions / The actual giving of the test</i>	

5. Do you believe the cut scores set by the panel are correctly placed on the assessment score scale?

- A. Definitely yes (1)
B. Probably yes (6)
C. Unsure (0)
D. Probably no (0)
E. Definitely no (0)
One respondent did not answer this question.

Please explain your answer:

- We are very close and we seemed to agree on most items*
- Yes, scores were discussed with consensus*
- We were very close in our assessment scores and I felt we were all on the same page*
- Our tables were very close on our cut scores across the board*
- The group was close in scores*
- Agreeing on additional details to the performance level descriptors was vital to the whole process*

6. How could the standard-setting process have been improved?

- This has been a great experience. I think most of the problems are going to be found in the actual testing. There are too many factors that will influence the scores*
- Well done, excellent moderator*
- It seemed great. I would be interested seeing other methods*
- We needed to use different systems for numbering. Everything seemed to use 0, 1, 2, 3, 4 and it confused the choices with using prompts on the test and the general answers*
- I think if you would use the same people to do this again for changes because everyone now is familiar with the process and it will become more accurate if it needs to be done*

For each statement below, please circle the rating that best represents your judgment.

7. The opening session was:

Not at All Useful				Very Useful
1 (0)	2 (1)	3 (3)	4 (3)	5 (1)

8. The performance level descriptors were:

Not at All Clear				Very Clear
1 (0)	2 (0)	3 (2)	4 (6)	5 (0)

9. Providing additional details to the performance level descriptors was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (0)	4 (3)	5 (5)

10. The discussion with other panelists was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (0)	4 (3)	5 (5)

11. The student profile rating task was:

Not at All Clear				Very Clear
1 (0)	2 (0)	3 (0)	4 (4)	5 (4)

12. The impact data at the beginning of round 3 was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (0)	4 (2)	5 (6)

Additional Comments

13. Please provide any additional comments or suggestions about the standard-setting process.

- *I hope you will seriously consider our "parking lot" suggestions and not just "file them away."*
- *I feel that the way we started by going through the ratings and agree on exactly what we had made the process much easier. I also feel that another category is needed for non-notable or inconclusive.*

MATH GRADE 10

Evaluation of the Standard Setting Procedures for the Montana Alternate Assessment

1. What is your overall impression of the process used to set performance standards for the Montana Alternate Assessment? (Circle one.)

- A. Very good (1)
- B. Good (6)
- C. Unsure (0)
- D. Poor (0)
- E. Very poor (0)

2. How clear were you with the performance level descriptors? (Circle one.)

- A. Very clear (5)
- B. Clear (3)
- C. Somewhat clear (1)
- D. Not clear (0)

3. How would you judge the length of time of this meeting for setting performance standards? (Circle one.)

- A. About right (9)
- B. Too little time (0)
- C. Too much time (0)

4. What factors influenced the standards you set? (For each, circle the most appropriate rating from 1=Not at all Influential to 5=Very Influential.)

- A. The performance level descriptors

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (0)	3 (0)	4 (4)	5 (5)

- B. The assessment items (performance indicators)

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (0)	3 (1)	4 (5)	5 (3)

- C. Other panelists

Not at All Influential		Moderately Influential		Very Influential
1 (0)	2 (0)	3 (3)	4 (5)	5 (1)

D. My experience in the field

Not at All Influential	Moderately Influential		Very Influential
1 (0)	2 (0)	3 (2)	4 (6)
			5 (1)

E. Other (please specify_____)

Not at All Influential	Moderately Influential		Very Influential
1 (0)			
2 (0)			
3 (1)	<i>Numbers of students at each level</i>		
4 (0)			
5 (0)			

5. Do you believe the cut scores set by the panel are correctly placed on the assessment score scale?

- A. Definitely yes (5)
 - B. Probably yes (3)
 - C. Unsure (0)
 - D. Probably no (0)
 - E. Definitely no (0)
- One respondent did not answer this question.*

Please explain your answer:

- *I disagreed on a few items, but we were all close*
- *Knowing the % age of students seemed important in making cut scores. It fell into place better once we knew % age of students represented by test scores*
- *Utilizing the data given, they are correct. If other data, such as numbers of participants for percentages could impact*
- *We had extensive discussion, shared valuable ideas*
- *We came to a good consensus after our group discussion*
- *As we discussed our reasoning for cuts, I feel we used the performance descriptors accurately in our decision-making process*
- *After the group discussion and data analysis, I felt very comfortable with decision*
- *Using the available information and the process, I believe the cut scores are well placed*

6. How could the standard-setting process have been improved?

- *Not sure*
- *I feel that the process was very helpful and that as a group, we worked very collaboratively and supported one another in a professional way*
- *Knowing the number of students in each group would have helped*
- *More accurate data, the uncertainty of blanks versus 0's, number of students at each level, etc.*

- *Loved the diverse backgrounds of the participants. This group was excellent for input from all. If all groups could have participants with comparable backgrounds*

For each statement below, please circle the rating that best represents your judgment.

7. The opening session was:

Not at All Useful				Very Useful
1 (0)	2 (1)	3 (0)	4 (5)	5 (2)

8. The performance level descriptors were:

Not at All Clear				Very Clear
1 (0)	2 (0)	3 (2)	4 (3)	5 (5)

9. Providing additional details to the performance level descriptors was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (0)	4 (1)	5 (8)

10. The discussion with other panelists was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (0)	4 (0)	5 (9)

11. The student profile rating task was:

Not at all Clear				Very Clear
1 (0)	2 (0)	3 (1)	4 (7)	5 (1)

12. The impact data at the beginning of round 3 was:

Not at All Useful				Very Useful
1 (0)	2 (0)	3 (0)	4 (5)	5 (4)

Additional Comments

13. Please provide any additional comments or suggestions about the standard-setting process.

- *The rooms were very nice. It was good to be closer to downtown. The diversity of the groups was good.*
- *Very well organized workshop!*
- *Having diverse backgrounds among participants was critically important! The different perspectives and orientations – extremely helpful*
- *The variety of people/backgrounds was very helpful – it provided different views that should be considered*
- *The diversity of participants is critical. Our facilitator did an excellent job*

Appendix D CRT-ALTERNATE PERFORMANCE LEVEL DESCRIPTORS

Advanced	The student at the Advanced level accurately and independently demonstrates the ability to carry out comprehensive content specific performance indicators.
Proficient	The student at the Proficient level, given limited prompting , demonstrates the ability to respond accurately in performing a wide variety of content specific performance indicators.
Nearing Proficiency	The student at the Nearing Proficiency level, given moderate prompting , demonstrates the ability to respond accurately in performing a narrow set of content specific performance indicators.
Novice	The student at the Novice level, given physical assistance and/or modeling , is supported to participate in content specific performance indicators.

CRT-Alternate Scaled Score Ranges for Performance Levels

Grade 4

	Reading	Mathematics
Advanced	258-300	300-300
Proficient	250-257	250-299
Nearing Proficiency	225-249	225-249
Novice	200-224	200-224

Grade 8

	Reading	Mathematics
Advanced	257-300	272-300
Proficient	250-256	250-271
Nearing Proficiency	225-249	225-249
Novice	200-224	200-224

Grade 10

	Reading	Mathematics
Advanced	265-300	283-300
Proficient	250-264	250-282
Nearing Proficiency	225-249	225-249
Novice	200-224	200-224

CRT-Alternate Cut Scores for Performance Levels

Raw-to-Scaled Score Correspondence Grade 4		
Raw Score	Reading Scaled Score	Math Scaled Score
0	206	200
1	207	200
2	208	200
3	208	200
4	209	200
5	209	200
6	210	200
7	211	200
8	211	200
9	212	200
10	213	200
11	213	200
12	214	200
13	214	200
14	215	200
15	216	200
16	216	200
17	217	200
18	218	200
19	218	200
20	219	200
21	219	200
22	220	200
23	221	200
24	221	200
25	222	200
26	223	200
27	223	200
28	224	200
29	224	200
30	225	200
31	226	200
32	226	200
33	227	200
34	228	200
35	228	200
36	229	200
37	229	200
38	230	200
39	231	200
40	231	200

41	232	200
42	233	200
43	233	200
44	234	200
45	234	200
46	235	200
47	236	200
48	236	200
49	237	200
50	238	200
51	238	200
52	239	200
53	239	200
54	240	202
55	241	204
56	241	206
57	242	208
58	243	210
59	243	213
60	244	215
61	244	217
62	245	219
63	246	221
64	246	223
65	247	225
66	248	227
67	248	229
68	249	231
69	249	233
70	250	235
71	251	238
72	251	240
73	252	242
74	253	244
75	253	246
76	254	248
77	254	250
78	255	252
79	256	254
80	256	256
81	257	258
82	257	260
83	258	263
84	259	265
85	259	267
86	260	269
87	261	271
88	261	273
89		275

90		277
91		279
92		281
93		283
94		285
95		288
96		290
97		292
98		294
99		296
100		298
101		299
102		299
103		299
104		299
105		300
106		300
107		300
108		300
109		300
110		300
111		300
112		300

Raw-to-Scaled Score Correspondence Grade 8		
Raw Score	Reading Scaled Score	Math Scaled Score
0	215	200
1	216	200
2	216	200
3	217	200
4	217	200
5	218	200
6	218	200
7	218	200
8	219	200
9	219	200
10	220	200
11	220	200
12	221	200
13	221	200
14	221	200
15	222	200
16	222	200
17	223	200

18	223	200
19	224	200
20	224	200
21	224	200
22	225	200
23	225	200
24	226	200
25	226	200
26	227	200
27	227	200
28	228	200
29	228	200
30	229	200
31	229	200
32	229	200
33	230	200
34	230	200
35	231	200
36	231	201
37	232	202
38	232	203
39	232	203
40	233	204
41	233	205
42	234	206
43	234	207
44	235	208
45	235	209
46	236	209
47	236	210
48	236	211
49	237	212
50	237	213
51	238	214
52	238	215
53	239	216
54	239	216
55	239	217
56	240	218
57	240	219
58	241	220
59	241	221
60	242	222
61	242	222
62	243	223
63	243	224
64	243	225
65	244	226
66	244	227

67	245	228
68	245	228
69	246	229
70	246	230
71	246	231
72	247	232
73	247	233
74	248	234
75	248	234
76	249	235
77	249	236
78	249	237
79	250	238
80	250	239
81	251	240
82	251	241
83	252	241
84	252	242
85	253	243
86	253	244
87	254	245
88	254	246
89	254	247
90	255	247
91	255	248
92	256	249
93	256	250
94	257	251
95	257	252
96	257	253
97		253
98		254
99		255
100		256
101		257
102		258
103		259
104		259
105		260
106		261
107		262
108		263
109		264
110		265
111		266
112		266
113		267
114		268
115		269

116		270
117		271
118		271
119		272
120		273
121		274
122		275
123		276
124		277
125		278
126		278
127		279
128		280

Raw-to-Scaled Score Correspondence Grade 10		
Raw Score	Reading Scaled Score	Math Scaled Score
0	200	200
1	200	200
2	200	200
3	200	200
4	200	200
5	200	200
6	200	200
7	200	200
8	200	200
9	200	200
10	200	200
11	200	200
12	200	200
13	200	200
14	200	200
15	200	200
16	200	200
17	200	200
18	200	200
19	200	200
20	200	200
21	200	200
22	200	200
23	200	200
24	200	200
25	200	200
26	200	200
27	200	200

28	200	200
29	200	200
30	200	200
31	200	200
32	200	200
33	200	200
34	200	200
35	200	200
36	200	200
37	200	200
38	200	200
39	200	200
40	200	200
41	200	200
42	200	200
43	200	200
44	200	200
45	200	200
46	200	201
47	200	202
48	200	203
49	200	205
50	200	206
51	200	207
52	200	208
53	200	209
54	200	210
55	201	211
56	203	212
57	204	214
58	205	215
59	206	216
60	208	217
61	209	218
62	210	219
63	211	220
64	213	222
65	214	223
66	215	224
67	216	225
68	218	226
69	219	227
70	220	228
71	221	230
72	223	231
73	224	232
74	225	233
75	226	234
76	228	235

77	229	236
78	230	238
79	231	239
80	233	240
81	234	241
82	235	242
83	236	243
84	238	244
85	239	245
86	240	247
87	241	248
88	243	249
89	244	250
90	245	251
91	246	252
92	248	253
93	249	255
94	250	256
95	251	257
96	253	258
97	254	259
98	255	260
99	256	261
100	258	263
101	259	264
102	260	265
103	261	266
104	263	267
105	264	268
106	265	269
107	266	270
108	268	272
109		273
110		274
111		275
112		276
113		277
114		278
115		280
116		281
117		282
118		283
119		284
120		285
121		286
122		288
123		289
124		290

Appendix E

CRT-Alternate Released Performance Indicators Montana CRT-Alternate Assessment 2005 Released Items (Performance Indicators) Reading Grade 4

Item	Performance Indicator	Standard
1	Responds to own name presented via any communicative modality.	1
2	Anticipates the beginning of a literacy activity.	1
3	Attends to another person demonstrating a procedure.	1
4	Attends to literacy materials.	1
5	Locate a picture/symbol/object when named or signed.	1
6	Responds to yes/no questions about information in print and nonprint materials.	5
7	Previews/explores resource materials.	1
8	Identifies appropriate resource to gain specific information.	1
9	Selects literacy materials/book by character.	1
10	Displays knowledge of front/back, right side up, page turning, and scanning when exploring literacy materials.	2
11	Uses auditory, visual, or tactile scanning to maintain place and follow along.	2
12	Identifies a word/picture/symbol/object that is new and unfamiliar.	2
13	Attends to literacy materials from beginning to end.	1
14	Identifies plot.	2
15	Demonstrate understanding of a word based on the context of a reading selection.	2
16	Answers "who" question about character in the story (using spoken words, pictures/symbols/objects or communication devices).	2
17	Answers "what" question about event or object in story.	2
18	Identifies supporting details from an expository reading/literary selection.	2
19	Answers "where" question about the place in a story.	2
20	Provides details about perspective.	4
21	Identifies events from a functional text.	4
22	Uses a timeline to provide information about an event.	4

Montana Standards for Reading:

Standard 1 – Students construct meaning as they comprehend, interpret, and respond to what they read.

Standard 2 – Students apply a range of skills and strategies to read.

Standard 3 – Students set goals, monitor, and evaluate their progress in reading.

Standard 4 – Students select, read, and respond to print and nonprint material for a variety of purposes.

Standard 5 – Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences.

**Montana CRT-Alternate Assessment
2005 Released Items (Performance Indicators)
Math Grade 4**

Item	Performance Indicator	Standard
1	Attends to another person demonstrating a procedure.	1
2	Anticipates the beginning of a math activity.	1
3	Attends to materials being displayed.	1
4	Demonstrates the concept of one.	2
5	Applies a number/word to a quantity of objects in a collection (few/many; one/many, more/less).	2
6	Determines which number is closer to the quantity in a given set.	2
		1
7	Counts using a sequential order of numbers.	2
8	Demonstrates one-to-one correspondence among up to 12 objects and counting numbers (rational counting).	2
9	Uses final number as quantity of a set.	2
10	Sorts objects into categories.	6
11	Represents data.	6
12	Sets up graph (i.e., labels axes).	6
13	Sets up graph (i.e., labels axes).	6
14	Makes a bar graph.	6
15	Finds the category with the most/least.	6
16	Answers questions about a graph.	6
		1
17	Compares categories.	6
		1
18	Communicates relationships between categories.	6
		1
		1
19	Describes or recognizes characteristics of categories.	6
		1
20	Communicates relationships between categories.	6
		1
21	Computes with addition.	2
22	Computes with subtraction.	2
23	Predicts outcome of a chance event.	6
24	Explains reasoning about probability problems.	6
25	Extends an alternating pattern.	7
26	Creates a repeating pattern.	7
27	Extends a pattern.	7
28	Creates a pattern.	7

Montana Standards for Mathematics:

Standard 1 – Students engage in the mathematical processes of problem solving and reasoning, estimation, communication, connections and applications, and using appropriate technology.

Standard 2 – Students demonstrate understanding of and an ability to use numbers and operations.

Standard 3 – Students use algebraic concepts, processes, and language to model and solve a variety of real-world and mathematical problems.

Standard 4 – Students demonstrate understanding of shape and an ability to use geometry.

Standard 5 – Students demonstrate understanding of measurable attributes and an ability to use measurement processes.

Standard 6 – The students demonstrate understanding of an ability to use data analysis, probability, and statistics.

Standard 7 – Students demonstrate understanding of an ability to use patterns, relations and functions.

**Montana CRT-Alternate Assessment
2005 Released Items (Performance Indicators)
Reading Grade 8**

Item	Performance Indicator	Standard
1	Responds to own name presented via any communicative modality.	1
2	Anticipates the beginning of a literacy activity.	1
3	Anticipates routines or patterns connected to literacy activity.	1
4	Identifies a word/picture/symbol/object used to name a familiar place.	2
5	Locates the library (reference area, or media center.	2
6	Identifies words/pictures/symbols/objects used to name familiar places.	2
7	Attends to literacy materials.	1
8	Indicates preference when offered a choice of books/materials.	1
9	Indicates adaptations needed to understand text.	1
10	Identifies resource materials to gain information about words.	1
11	Displays knowledge of front/back, right side up, page turning, and scanning when exploring literacy materials.	2
12	Selects literacy materials/books by character or topic.	1
13	Recalls name of common object/symbol when given the function of the object.	1
14	Identifies words/pictures/symbols/objects to name familiar people.	2
15	Identifies a word/picture/symbol/object used for content communication.	1
16	Identifies the appropriate resource to gain specific information.	4
17	Locates titles.	2
18	Identifies letters by name/signing.	2
19	Recognizes vowel letter-sound association.	2
20	Indicates that a sentence is made up of words.	2
21	Uses auditory or visual scanning to maintain place.	2
22	Identifies facts in text.	5
23	Distinguishes fact from opinion.	5
24	Identifies, locates, reads and interprets information from a variety of documents and sources.	4

Montana Standards for Reading:

Standard 1 – Students construct meaning as they comprehend, interpret, and respond to what they read.

Standard 2 – Students apply a range of skills and strategies to read.

Standard 3 – Students set goals, monitor, and evaluate their progress in reading.

Standard 4 – Students select, read, and respond to print and nonprint material for a variety of purposes.

Standard 5 – Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences.

**Montana CRT-Alternate Assessment
2005 Released Items (Performance Indicators)
Mathematics Grade 8**

Item	Performance Indicator	Standard
1	Attends to another person demonstrating a procedure.	1
2	Anticipates the beginning of a math activity.	1
3	Attends to materials being displayed.	1
4	Attends to another person showing relationship between two variables, using objects, picture, symbols, or numbers.	1
5	Demonstrates the concept of "one."	2
6	Determines questions for obtaining data.	6
7	Describes features of the data.	6
8	Counts with another person.	2
9	Creates a frequency table.	6
10	Creates a simple graph/frequency plot using real objects and/or symbols.	6
11	Displays two or more categories on a bar graph.	6
12	Sets up a graph; labels axes.	6
13	Sets up a graph; labels axes.	6
14	Explains how to use a bar graph.	6
15	Determines which category has the most/least votes.	6
16	Uses tables or graphs to make decisions.	6 1
17	Shows a quantity.	2
18	Demonstrates understanding of some/more/less.	2
19	Computes an addition problem.	2
20	Supplies the missing number represented by a blank in a number sentence in which the operation might be addition, subtraction, or multiplication.	3
21	Supplies the missing number represented by a blank in a number sentence in which the operation might be addition, subtraction, or multiplication.	3
22	Shows a relationship between two variables.	3
23	Given a numerical relationship between two variables and the value of one variable, finds the other.	3
24	Uses a table to make decisions.	6 1
25	Uses ruler to measure objects that are a whole number of inches long.	5
26	Measures with a ruler.	5
27	Uses an appropriate unit of measure.	5
28	Demonstrates reasoning to solve a measurement problem.	5
29	Measures with a yardstick.	5
30	Uses a calculator for computation.	2
31	Chooses a correct procedure to solve a problem.	2 1
32	Produces fractional parts of a whole.	2

Montana Standards for Mathematics:

Standard 1 – Students engage in the mathematical processes of problem solving and reasoning, estimation, communication, connections and applications, and using appropriate technology.

Standard 2 – Students demonstrate understanding of and an ability to use numbers and operations.

Standard 3 – Students use algebraic concepts, processes, and language to model and solve a variety of real-world and mathematical problems.

Standard 4 – Students demonstrate understanding of shape and an ability to use geometry.

Standard 5 – Students demonstrate understanding of measurable attributes and an ability to use measurement processes.

Standard 6 – The students demonstrate understanding of an ability to use data analysis, probability, and statistics.

Standard 7 – Students demonstrate understanding of an ability to use patterns, relations and functions.

**Montana CRT-Alternate Assessment
2005 Released Items (Performance Indicators)
Reading Grade 10**

Item	Performance Indicator	Standard
1	Responds to own name presented via any communicative modality.	1
2	Anticipates the beginning of a literacy activity.	1
3	Attends to another person demonstrating a procedure.	1
4	Attends to literacy materials.	2
5	Previews/explores reading materials.	1
6	Locates picture/object/symbol when named or signed.	1
7	Identifies a variety of resources.	1
8	Demonstrates understanding of difference between information and literature.	4
9	Demonstrates an understanding/awareness of prior knowledge of concept.	1
10	Identifies appropriate information resource to gain specific information.	4
11	Selects literacy materials/books by character or topic.	1
12	Indicates preference when offered a choice of books.	1
13	Identifies word/picture/symbol/object used for content communication.	1
14	Uses text features to move through text in appropriate sequence.	2
15	Follows directions that contain verbs.	1
16	Communicates an opinion.	1
17	Identifies words/pictures/symbols/objects used to name familiar people.	2
18	Displays knowledge of front/back, right side up, page turning, scanning, when exploring literacy material.	2
19	Uses auditory or visual scanning to maintain place.	2
20	Identifies the main idea of expository reading selection.	1
21	Uses a picture/object to identify activity or item.	2
22	Uses one source to organize information	5
23	Uses pictures/symbols/objects to communicate abstract meaning.	2
24	Uses graphic organizers to identify similarity and difference.	5
25	Identifies time of events from a schedule.	4
26	Uses graphic organizers to identify similarities and differences.	5
27	Makes connections, explain relationships among a variety of sources, and integrate similar information.	5

Montana Standards for Reading:

Standard 1 – Students construct meaning as they comprehend, interpret, and respond to what they read.

Standard 2 – Students apply a range of skills and strategies to read.

Standard 3 – Students set goals, monitor, and evaluate their progress in reading.

Standard 4 – Students select, read, and respond to print and nonprint material for a variety of purposes.

Standard 5 – Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences.

**Montana CRT-Alternate Assessment
2005 Released Items (Performance Indicators)
Math Grade 10**

Item	Performance Indicator	Standard
1	Anticipates the beginning of a math activity.	1
2	Attends to materials being displayed.	1
3	Attends to another person showing relationship between two variables, using objects, picture, symbols, or numbers.	1
4	Attends to another person demonstrating with concrete materials.	1
5	Demonstrates that a collection of objects has a quantity.	2
6	Demonstrates the concept of one.	2
7	Matches bills and their values.	2
8	Uses different bill combinations to show equivalent amounts.	2
9	Uses different bill combinations to show equivalent amounts.	2
10	Demonstrates an understanding of multiplication using concrete materials.	2
11	Demonstrates an understanding of multiplication and division through concrete materials.	2
12	Recognizes properties of 2-dimensional shapes.	4
13	Identifies a square regardless of its orientation, in general shape.	4
14	Follows navigational directions.	4
15	Covers a figure with shapes.	4
16	Chooses addition.	2
17	Chooses correct strategies or procedures to solve a number problem.	2 1
18	Attends to another person showing relationship between two variables.	3
19	Given a numerical relationship between two variables and the value of one of the variables, finds the other variable.	3
20	Uses or extends a T-table to find value of a variable.	3
21	Uses or extends a T-table to find value of a variable.	3
22	Uses or extends a T-table to find value of a variable.	3
23	Determines change.	2
24	Determines how much more money is needed.	2
25	Attends to another person demonstrating with concrete materials.	7
26	Models mathematical problems.	7
27	Computes addition problems.	2
28	Computes addition problems.	2
29	Shows relationship between two variables.	3
30	Given a mathematical relationship between two variables and the value of one variable, finds the value of the other variable.	3
31	Uses tables to make decisions.	7

Montana Standards for Mathematics:

Standard 1 – Students engage in the mathematical processes of problem solving and reasoning, estimation, communication, connections and applications, and using appropriate technology.

Standard 2 – Students demonstrate understanding of and an ability to use numbers and operations.

Standard 3 – Students use algebraic concepts, processes, and language to model and solve a variety of real-world and mathematical problems.

Standard 4 – Students demonstrate understanding of shape and an ability to use geometry.

Standard 5 – Students demonstrate understanding of measurable attributes and an ability to use measurement processes.

Standard 6 – The students demonstrate understanding of an ability to use data analysis, probability, and statistics.

Standard 7 – Students demonstrate understanding of an ability to use patterns, relations and functions.

Appendix F Report Shells

Student Report

Class Roster & Item-Level Report

School Summary Report

System Summary Report

CRT-Alternate Performance Level Descriptors

The Performance Level Descriptors below describe students' knowledge, skills, and abilities in a content area. These descriptions provide a picture or profile of student achievement at the four performance levels: **Advanced**, **Proficient**, **Nearing Proficiency**, and **Novice**.

Advanced

The student at the Advanced level accurately and independently demonstrates the ability to carry out comprehensive content specific performance indicators.

Proficient

The student at the Proficient level, given limited prompting, demonstrates the ability to respond accurately in performing a wide variety of content specific performance indicators.

Nearing Proficiency

The student at the Nearing Proficiency level, given moderate prompting, demonstrates the ability to respond accurately in performing a narrow set of content specific performance indicators.

Novice

The student at the Novice level, given physical assistance and/or modeling, is supported to participate in content specific performance indicators.

Score Ranges

	Reading	Math
Advanced	(258-300)	(300-300)
Proficient	(250-257)	(250-299)
Nearing Proficiency	(225-249)	(225-249)
Novice	(200-224)	(200-224)

For more information regarding student assessments in Montana, check out the Office of Public Instruction's Parents Page at www.opi.state.mt.us/parents/.

OPI Contact
Judy Snow, State Assessment Director
406-444-3656
jsnow@mt.gov



Linda McCulloch, Superintendent
Montana Office of Public Instruction
PO Box 202501
Helena, Montana 59620-2501
<http://www.opi.mt.gov>

Criterion-Referenced Test (CRT-Alternate) MontCAS, Phase 2 Student Report 2005

Student Name:
School:
System:
Grade: 04

Dear Parents/Guardians:

This report contains the results of the second year of the Montana Comprehensive Assessment System Criterion-Referenced Test-Alternate (CRT-Alternate) that your child took in February and March. The major purpose of the CRT-Alternate is to provide schools with solid information to evaluate and improve curriculum and instruction to help all students meet Montana's reading and mathematics standards. This report provides important information about your child's performance on the assessment, along with state results.

Your child participated in the CRT-Alternate Assessment. The CRT-Alternate measures your child's performance based on alternate achievement standards. The CRT-Alternate is aligned with the Montana State Standards for Reading and Mathematics. Test results are based on teacher observation of your child's performance on specifically designed tasks. Your child's results in reading and mathematics are reported in one of four performance levels. The performance levels are defined on the back cover of this report.

It is important to remember that the CRT-Alternate is just one measure of your child's academic progress. Your local school staff can provide further information about your child's performance in school. The CRT-Alternate, which is required by the No Child Left Behind Act, is part of an ongoing statewide educational improvement process. Working together, we can ensure that Montana's children continue to receive a high-quality education.

Sincerely,

Linda McCulloch
Montana Superintendent of Public Instruction

How did YOUR CHILD do on the CRT-Alternate?

Scaled Scores on the CRT-Alternate

The criterion-referenced test-alternate (CRT-Alternate) is designed to measure student performance against the learning goals described in the Montana Content Standards (<http://www.opi.state.mt.us/standards/index.html>). Consistent with this purpose, results on the CRT-Alternate are reported according to performance levels that describe student performance in relation to the established state standards. There are four performance levels: **Advanced**, **Proficient**, **Nearing Proficiency**, and **Novice**. Your child's performance levels in reading and mathematics are based on a total scaled score in each content area. Scaled scores in each content area range from 200 to 300. Your child's performance levels, based on the scaled scores, are shown in the bar graphs below.

Scaled Scores

STUDENT RESULTS FOR READING

Performance Level:
Student Scaled Score:



STUDENT RESULTS FOR MATHEMATICS

Performance Level:
Student Scaled Score:



Scores on Montana Content Standards

In addition to performance levels, CRT-Alternate results are reported for Montana Content Standards in reading and mathematics. Unlike scaled scores which provide a total performance level score, Montana Content Standard Scores provide more specific information about your child's achievement on the CRT-Alternate. The chart on the following page shows your child's performance in each area of study within subject areas (Montana Content Standards for reading and math). These results can be used to show your child's relative strengths or weaknesses.

Contact your student's school for more information about the following symbols:

† Student did not complete the assessment.

§ Teacher halted the administration of the assessment after the student scored a 0 for three consecutive items on two different test administrations.

Scores on Montana Standards

Percentage of Points Earned

Reading Standards	Points Possible	Student Percentage	State Percentage	Percentage of Points Earned				
				0	25	50	75	100
1. Students construct meaning as they comprehend, interpret, and respond to what they read.	36							
2. Students apply a range of skills and strategies to read.	36							
3. Students set goals, monitor, and evaluate their reading progress.	This standard is not measureable in a statewide assessment.							
4. Students select, read, and respond to print and nonprint material for a variety of purposes.	12							
5. Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences.	4	—	—					
Math Standards								
1. Problem Solving	36							
2. Numbers and Operations	32							
3. Algebra	0	—	—					
4. Geometry	0	—	—					
5. Measurement	0	—	—					
6. Data Analysis, Statistics, and Probability	52							
7. Patterns, Relations, and Functions	16							

Percentage of points earned by student

State percentage of points earned

—There were too few score points to report on this standard, or no items on the test measured this standard.

**MontCAS, Phase 2
CRT-Alternate**

**Reading
Roster & Item-Level Report
Confidential**

Class:
School:
System:

Spring 2005

Grade: 04
Page: 1 of: 1

[illegible]

† Student did not complete the assessment. ‡ Teacher halted the administration of the assessment after the student scored a 0 for three consecutive items on two different test administrations. § Not in school and/or district for full academic year.

ShellsShells

MontCAS, Phase 2 CRT-Alternate

Reading

School Summary Report

School:
System:
Grade: 04
Spring 2005

I. Distribution of scores

Perf. Level	Scores	School			System			State		
		N	% of Students	% of Students in Cal.	N	% of Students	% of Students in Cal.	N	% of Students	% of Students in Cal.
Advanced	292-300									
	284-291									
	275-283									
	267-274									
	258-266									
Proficient	256-257									
	255-255									
	253-254									
	252-252									
	250-251									
Near Proficiency	245-249									
	240-244									
	235-239									
	230-234									
	225-229									
Novice	220-224									
	215-219									
	210-214									
	205-209									
	200-204									

*The sum of the points for each standard may exceed the total points, as some items correlate with more than one standard.

II. Subtest results

Reading	Points Possible	Average Points Earned		
		School	System	State
*Total Points	88			
1. Students construct meaning as they comprehend, interpret, and respond to what they read	36			
2. Students apply a range of skills and strategies to read	36			
3. Students set goals, monitor, and evaluate their reading progress	This standard is not measureable in a statewide assessment.			
4. Students select, read, and respond to print and nonprint material for a variety of purposes	12			
5. Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences	4	—	—	—

—There were too few score points to report on this standard, or no items on the test measured this standard.

CRT-Alternate Performance Level Descriptors

Advanced (258-300)

The student at the Advanced level accurately and independently demonstrates the ability to carry out comprehensive content specific performance indicators.

Proficient (250-257)

The student at the Proficient level, given limited prompting, demonstrates the ability to respond accurately in performing a wide variety of content specific performance indicators.

Nearing Proficiency (225-249)

The student at the Nearing Proficiency level, given moderate prompting, demonstrates the ability to respond accurately in performing a narrow set of content specific performance indicators.

Novice (200-224)

The student at the Novice level, given physical assistance and/or modeling, is supported to participate in content specific performance indicators.

ShellsShells

MontCAS, Phase 2 CRT-Alternate

Confidential

Reading

School
Summary
Report

School:
System:
Grade: 04
Spring 2005

III. Results for Subgroups of Students

Reporting category	School					System					State				
	N	% in N	% in NP	% in P	% in A	N	% in N	% in NP	% in P	% in A	N	% in N	% in NP	% in P	% in A
All Students															
Gender															
Male															
Female															
Ethnicity															
American Indian															
Asian															
Hispanic															
Black or African American															
Native Hawaiian or Other Pacific Islander															
White															
Significant Cognitive Disability															
Special Education															
Students with a 504 Plan															
Title I (optional)															
Tested with Standard Accommodation															
Tested with Non-Standard Accommodation															
Alternate Assessment															
Migrant															
Gifted/Talented															
LEP/ELL															
Former LEP Student															
LEP Student Enrolled for First Time in a U.S. School		Performance levels are not reported for 1st year LEP students.													
Free/Reduced Lunch															
Special Education (Disability):															
Autism															
Child with a Disability															
Cognitive Delay															
Deaf-Blindness Impairment															
Deafness															
Emotional Disturbance															
Hearing Impairment															
Learning Disability															
Orthopedic Impairment															
Other Health Impairment															
Speech/Language															
Traumatic Brain Injury															
Visual Impairment															

*Less than ten (10) students were assessed.

ShellsShells

MontCAS, Phase 2 CRT-Alternate

System:
Grade: 04
Spring 2005

Reading

System Summary Report

I. Distribution of scores

Perf. Level	Scores	System			State		
		N	% of Students	% of Students in Cat.	N	% of Students	% of Students in Cat.
Advanced	292-300						
	284-291						
	275-283						
	267-274						
	258-266						
Proficient	256-257						
	255-255						
	253-254						
	252-252						
	250-251						
Nearing Proficiency	245-249						
	240-244						
	235-239						
	230-234						
	225-229						
Novice	220-224						
	215-219						
	210-214						
	205-209						
	200-204						

* The sum of the points for each standard may exceed the total points, as some items correlate with more than one standard.

II. Subtest results

Reading		Points Possible	Average Points Earned	
			System	State
* Total Points		88		
Standards	1. Students construct meaning as they comprehend, interpret, and respond to what they read	36		
	2. Students apply a range of skills and strategies to read	36		
	3. Students set goals, monitor, and evaluate their reading progress	This standard is not measureable in a statewide assessment.		
	4. Students select, read, and respond to print and nonprint material for a variety of purposes	12		
	5. Students gather, analyze, synthesize, and evaluate information from a variety of sources, and communicate their findings in ways appropriate for their purposes and audiences	4	—	—

— There were too few score points to report on this standard, or no items on the test measured this standard.

CRT-Alternate Performance Level Descriptors

Advanced (258-300)

The student at the Advanced level accurately and independently demonstrates the ability to carry out comprehensive content specific performance indicators.

Proficient (250-257)

The student at the Proficient level, given limited prompting, demonstrates the ability to respond accurately in performing a wide variety of content specific performance indicators.

Nearing Proficiency (225-249)

The student at the Nearing Proficiency level, given moderate prompting, demonstrates the ability to respond accurately in performing a narrow set of content specific performance indicators.

Novice (200-224)

The student at the Novice level, given physical assistance and/or modeling, is supported to participate in content specific performance indicators.

Shells

MontCAS, Phase 2 CRT-Alternate

Confidential

Reading

System
Summary
Report

System:
Grade: 04
Spring 2005

III. Results for Subgroups of Students

Reporting category	System					State				
	N	% in N	% in NP	% in P	% in A	N	% in N	% in NP	% in P	% in A
All Students										
Gender										
Male										
Female										
Ethnicity										
American Indian										
Asian										
Hispanic										
Black or African American										
Native Hawaiian or Other Pacific Islander										
White										
Significant Cognitive Disability										
Special Education										
Students with a 504 Plan										
Title I (optional)										
Tested with Standard Accommodation										
Tested with Non-Standard Accommodation										
Alternate Assessment										
Migrant										
Gifted/Talented										
LEP/ELL										
Former LEP Student										
LEP Student Enrolled for First Time in a U.S. School		Performance levels are not reported for 1st year LEP students.								
Free/Reduced Lunch										
Special Education Disability(ies):										
Autism										
Child with a Disability										
Cognitive Delay										
Deaf-Blindness Impairment										
Deafness										
Emotional Disturbance										
Hearing Impairment										
Learning Disability										
Orthopedic Impairment										
Other Health Impairment										
Speech/Language										
Traumatic Brain Injury										
Visual Impairment										

*Less than ten (10) students were assessed.

Shellee

APPENDIX G:
Montana Alternate Assessment Spring 2005
FINAL Decision Rules

Participation			Relationship w/ Data file layout	Impact on Analyses	Impact on Student Report	Impact on School/System/State Report	Impact on Student Roster	Impact on student level data Excel files for system CDs	Impact on student level data Excel files for state CDs
Number of students("N ")	1	Number of students included in state aggregation.	NA	N=total number of students with 3 or more responses minus students tested at a private accredited school(PRAS) minus students tested in a non-accredited Title I school(PRNONST) minus students not enrolled(SNE) minus student enrolled part-time (PSNE) minus students tested at a private non-accredited school (PPNAC) minus LEP student					
No class header provided	2	No class indicators provided	Tfname=' ' and Tlname=' '	Class aggregations calculated are actually school level.	No impact	No impact	Report produced.	No impact	No impact
Number of Students for Reporting	3	Schools (Systems) has less than 10 included students in both content areas	NA		No impact	School/system report Produced. Page 2: For each category numbers will be suppressed if number of included students less than ten. The N-size is always reported	No Impact	No Impact	No Impact

Participation			Relationship w/ Data file layout	Impact on Analyses	Impact on Student Report	Impact on School/System/State Report	Impact on Student Roster	Impact on student level data Excel files for system CDs	Impact on student level data Excel files for state CDs
Tested but fewer than 3 items	4	Student responded to fewer than 3 items		Student not counted in N	Score given with a footnote (†) "Student did not complete the assessment"	Student not included	Score given with a footnote (†), "Student did not complete the assessment."	Student included	Student included
Program information	5	Student is identified as participating in an identified program.	Plan504='1' or Migrant='1' or LEP='1' or Lunch='1' or TM='1' or TR='1' or SE='1' or Disab='1'	If one or more Program Information codes are bubbled, student is counted as a program participant. LEP students do not include LEP students first time in U.S. school.	No impact	Reported on school & system reporting category reports. All numbers except the N-size are suppressed if N-size less than 10. Footnote * 'Less than 10 students were assessed.'	No impact	No impact	No impact
First year LEP student(not enrolled first time in U.S.school)	6	Student is identified as being a first year LEP student enrolled first time in U.S school.	LEPFirst not ='1' And Exclusions='1'	Student is excluded from all aggregations for both content areas.	Student receives report. Student does not receive scaled score or performance level for Reading. Performance Level='LEP' on report for reading. Student receives earned score in	Student is excluded from aggregations. Included in the count of 1 st yr LEP students on Page 2 of summary reports. Only N-size is reported. The rest of the line is covered with a watermark on Reading reports.	Student is included. Student's scaled score is blank. Student's performance level ='LEP' for Reading. If student took the Reading test the responses are shown. The student included in Math with	Student included	Student included
Student not enrolled (SNE) (Homescho oled)	7	Student is identified as not enrolled in an accredited public school.	Exclusions='3'	Student is not included in any school/system/state aggregations.	Student receives report	Not included on reports	Not included on reports	Students are not included on System CD	Not included on State CD

Participation			Relationship w/ Data file layout	Impact on Analyses	Impact on Student Report	Impact on School/System/State Report	Impact on Student Roster	Impact on student level data Excel files for system CDs	Impact on student level data Excel files for state CDs
Private Accredited School (PRAS)	8	Student is identified as testing at a private accredited public school.	Exclusions='5'	Student is not included in any state aggregations.	Student receives report	School report produced. System report produced. They are their own system.	Report produced	Students are included on System CD	Included on State CD; identified as PRAS
Private Non- Accredited Title 1 School(PRN ONST)	9	Student is identified as testing in a non- accredited Title 1 school	Exclusions='7'	Student is not included in any state aggregations.	Student receives report.	School report produced. System report produced. They are their own systems	Report produced	Students are included on System CD	Included on State CD; identified as PRNONST
Private Non- Accredited School (PRNAS)	10	Student is identified as testing in a non- accredited school	Exclusions='6'	Student is not included in state aggregations	Student receives report.	School report produced.	Report produced.	Students are included on System CD.	Included on state CD; identified as PRNAS
Student enrolled part- time(<180 hours)	11	Student is identified as enrolled part- time	Exclusions='4'	Student is not included in any school/system/state aggregations	Student receives report.	Student not included on school report.	Not included on reports.	Students are not included in system CD.	Student included on state CD; identified as PSNE

Participation			Relationship w/ Data file layout	Impact on Analyses	Impact on Student Report	Impact on School/System/State Report	Impact on Student Roster	Impact on student level data Excel files for system CDs	Impact on student level data Excel files for state CDs
Student did not complete the test	12	Teacher halted the administration of the assessment after the student scored 0 for three consecutive items. After three consecutive 0s blank out the rest of the scores.		Student receives earned score and performance level.	Student receives earned score and performance level with a footnote indicated by ('\$') "Teacher halted the administration of the assessment after the student scored a 0 for three consecutive items on two different test administrations"	Student included in aggregation.	Student receives earned score and performance level with a footnote indicated by ('\$') "Teacher halted the administration of the assessment after the student scored a 0 for three consecutive items on two different test administrations"	Student included in data file with a '1' in the halted field	Student included in data file with a '1' in the halted field.
Participation Information (NSAY & NDAY)	13	Student participated in CRT-Alternate but has not been a student in school or district for entire academic year	NA	Student is included in participation. If student is marked as NSAY only, then student is not included in school aggregations. If student is marked as NDAY, then student is not included in either school or district aggregations.	No impact.	If student is marked as NSAY only, then student is not included in school data. If student is marked as NDAY, then student is not included in school or district data.	If student is NSAY or NDAY student is included on roster with footnote(¥) "Not in school and/or district full academic year." Student excluded from school (if NSAY or NDAY) and/or district (if NDAY) aggregations.	No Impact	No Impact

Participation			Relationship w/ Data file layout	Impact on Analyses	Impact on Student Report	Impact on School/System/State Report	Impact on Student Roster	Impact on student level data Excel files for system CDs	Impact on student level data Excel files for state CDs
Special Education codes(not optional; can have more than one coded)	14	Student is has an identified disability under IDEA- 97.	AU='1',CW='1',CD ='1',DB='1',DE='1', ED='1',HI='1',LD=' 1',OI='1',OH='1', SL='1',TB='1',VI='1 ,	Student is counted in their respective disability group on page 2 of summary reports.	No Impact	Student is counted in their respective disability group on page 2 of summary reports. All numbers except the N-size are suppressed if N-size less than 10. Footnote * 'Less than 10 students were assessed.'	No Impact	No Impact	No Impact
Former LEP	15	Former LEP student	FLEP='1'	Student is included in all aggregations	Student receives report	Student included. Counted in category on page 2. All number except N-size are suppressed if N-size less than 10. Footnote * 'Less than 10 students were assessed.'	Student included	Student included	Student included
LEP currently receiving Title III services(not first year LEP)	16	LEP student currently receiving Title III services	Title3='1'	Student is included in all aggregations.	Student receives report	Student included. Counted in category on page 2.All numbers except N-size are suppressed if N-size less than 10. Footnote * 'Less than 10 students were assessed.'	Student included	Student included	Student included

Participation			Relationship w/ Data file layout	Impact on Analyses	Impact on Student Report	Impact on School/System/State Report	Impact on Student Roster	Impact on student level data Excel files for system CDs	Impact on student level data Excel files for state CDs
1 st year LEP enrolled first time in U.S school	17	LEP student enrolled for first time in U.S School	LEPFirst ='1'	Student is excluded from all aggregations for both content areas.	Student receives report. Student does not receive scaled score or performance level for Reading. Performance Level='LEP' on report for reading. Student receives earned score in Math.	Student is excluded from aggregations. Included in the count of 1 st yr LEP enrolled first time in U.S school students on Page 2 of summary reports. Only N-size is reported. The rest of the line is covered with a watermark on Reading reports.	Student is included. Student's scaled score is blank. Student's performance level ='LEP' for Reading. If student took the Reading test the responses are shown. The student included in Math with earned scores and responses shown.	Student included	Student included

Note:
Summary reports are generated for all participating schools regardless of number of included students.

Grade	Content	Total Possible Score
4	Reading	88
4	Math	112
8	Reading	96
8	Math	128
10	Reading	108
10	Math	124